

格致方法·定量研究系列　吴晓刚　主编

流动表分析

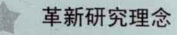
[美] 迈克尔·豪特（Michael Hout）著
李 骏 译

革新研究理念

丰富研究工具

最权威、最前沿的定量研究方法指南

17



格 致 出 版 社　上海人民出版社



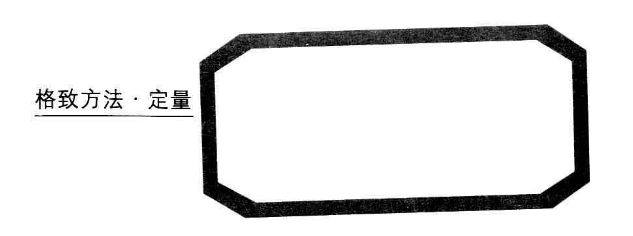
格致方法·定量研究系列

1. 社会统计的数学基础
2. 理解回归假设
3. 虚拟变量回归
4. 多元回归中的交互作用
5. 回归诊断简介
6. 现代稳健回归方法
7. 固定效应回归模型
8. 用面板数据做因果分析
9. 多层次模型
10. 分位数回归模型
11. 空间回归模型
12. 删截、选择性样本及截断数据的回归模型
13. 应用logistic回归分析（第二版）
14. logit与probit：次序模型和多类别模型
15. 定序因变量的logistic回归模型
16. 对数线性模型
17. 流动表分析
18. 关联模型
19. 中介作用分析
20. 因子分析：统计方法与应用问题
21. 非递归因果模型
22. 评估不平等
23. 分析复杂调查数据（第二版）
24. 分析重复调查数据
25. 世代分析
26. 纵贯研究
27. 多元时间序列模型
28. 潜变量增长曲线模型
29. 缺失数据
30. 社会网络分析
31. 广义线性模型导论
32. 基于行动者的模型
33. 基于布尔代数的比较法导论
34. 微分方程：一种建模方法
35. 模糊集合理论在社会科学中的应用
36. 图形代数
37. 项目功能差异

上架建议：社会研究方法
ISBN 978-7-5432-2126-



定价：15.00元
易文网：www.ewen.cc
格致网：www.hibooks.cn



流动表分析

[美] 迈克尔·豪特(Michael Hout) 著

李 骏 译

SAGE Publications ,Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

流动表分析/(美)豪特(Hout, M.)著;李骏译.
—上海:格致出版社:上海人民出版社,2012
(格致方法·定量研究系列)
ISBN 978-7-5432-2126-0

Ⅰ.①流… Ⅱ.①豪… ②李… Ⅲ.①定量社会学-研究方法 Ⅳ.①C91-03

中国版本图书馆CIP数据核字(2012)第131172号

责任编辑 高 璇

格致方法·定量研究系列
流动表分析
[美]迈克尔·豪特 著
李骏 译

出 版 世纪出版集团 www.ewen.cc
格致出版社 www.hibooks.cn
上海人民出版社
(200001 上海福建中路193号24层)

编辑部热线 021-63914988
市场部热线 021-63914081

发 行 世纪出版集团发行中心
印 刷 浙江临安曙光印务有限公司
开 本 920×1168毫米 1/32
印 张 5.5
字 数 103,000
版 次 2012年7月第1版
印 次 2012年7月第1次印刷
ISBN 978-7-5432-2126-0/C·76
定 价 15.00元

出版说明

由香港科技大学社会科学部吴晓刚教授主编的"格致方法·定量研究系列"丛书，精选了世界著名的SAGE出版社定量社会科学研究丛书中的35种，翻译成中文，集结成八册，于2011年出版。这八册书分别是：《线性回归分析基础》、《高级回归分析》、《广义线性模型》、《纵贯数据分析》、《因果关系模型》、《社会科学中的数理基础及应用》、《数据分析方法五种》和《列表数据分析》。这套丛书自出版以来，受到广大读者特别是年轻一代社会科学工作者的欢迎，他们针对丛书的内容和翻译都提出了很多中肯的建议。我们对此表示衷心的感谢。

基于读者的热烈反馈，同时也为了向广大读者提供更多的方便和选择，我们将该丛书以单行本的形式再次出版发行。在此过程中，主编和译者对已出版的书做了必要的修订和校正，还新增加了两个品种。此外，曾东林、许多多、范新光、李忠路协助主编参加了校订。今后我们将继续与SAGE出版社合作，陆续推出新的品种。我们希望本丛书单行本的出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

总序

往事如烟，光阴如梭。转眼间，出国已然十年有余。1996年赴美留学，最初选择的主攻方向是比较历史社会学，研究的兴趣是中国的制度变迁问题。以我以前在国内所受的学术训练，基本是看不上定量研究的。一方面，我们倾向于研究大问题，不喜欢纠缠于细枝末节。国内一位老师的话给我的印象很深，大致是说：如果你看到一堵墙就要倒了，还用得着纠缠于那堵墙的倾斜角度究竟是几度吗？所以，很多研究都是大而化之，只要说得通即可。另一方面，国内（十年前）的统计教学，总的来说与社会研究中的实际问题是相脱节的。结果是，很多原先对定量研究感兴趣的学生在学完统计之后，依旧无从下手，逐渐失去了对定量研究的兴趣。

我所就读的美国加州大学洛杉矶分校社会学系，在定量研究方面有着系统的博士训练课程。不论研究兴趣是定量还是定性的，所有的研究生第一年的头两个学期必须修两门中级统计课，最后一个学期的系列课程则是简单介绍线性回归以外的其他统计方法，是选修课。希望进一步学习定量研

究方法的可以在第二年修读另外一个三学期的系列课程，其中头两门课叫“调查数据分析”，第三门叫“研究设计”。除此以外，还有如“定类数据分析”、“人口学方法与技术”、“事件史分析”、“多层线性模型”等专门课程供学生选修。该学校的统计系、心理系、教育系、经济系也有一批蜚声国际的学者，提供不同的、更加专业化的课程供学生选修。2001 年完成博士学业之后，我又受安德鲁·梅隆基金会资助，在世界定量社会科学研究的重镇密歇根大学从事两年的博士后研究，其间旁听谢宇教授为博士生讲授的统计课程，并参与该校社会研究院(Institute for Social Research)定量社会研究方法项目的一些讨论会，受益良多。

2003 年，我赴港工作，在香港科技大学社会科学部，教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来，非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课，而第二门课“社会科学中的定量分析”为博士生的必修课(事实上，大部分硕士生在修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候，根据社会科学研究生的数理基础比较薄弱的特点，尽量避免复杂的数学公式推导，而用具体的例子，结合语言和图形，帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上，即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”，学完这些课程后，我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章；另一方面，也能在自己的研究中运用这些成熟的

方法论技术。

上述两门课的内容，尽管在线性回归模型的内容上有少量重复，但各有侧重。“社会科学里的统计学”(Statistics for Social Science)从介绍最基本的社会研究方法论和统计学原理开始，到多元线性回归模型结束，内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型，以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法，将重点放在因变量为定类数据的分析模型上，包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型，包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生，一直鼓励和支持我将两门课的讲稿结集出版，并帮助我将原来的英文课程讲稿译成了中文。但是，由于种种原因，这两本书拖了四年多还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩，每本书都写得通俗易懂。中山大学马骏教授向格致出版社何元龙社长推荐了这套书，当格致出版社向我提出从这套丛书中精选一批翻译，以飨中文读者时，我非常支持这个想法，因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种

语言的精准把握能力，还要有对实质内容有较深的理解能力，而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容，只有语言能力是远远不能胜任的。在短短的一年时间里，我们组织了来自中国内地及港台地区的二十几位研究生参与了这项工程，他们目前大部分是香港科技大学的硕士和博士研究生，受过严格的社会科学统计方法的训练，也有来自美国等地对定量研究感兴趣的博士研究生。他们是：

香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛，硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊，应用社会经济研究中心研究员李俊秀；香港大学教育学院博士研究生洪岩璧；北京大学社会学系博士研究生李丁、赵亮员；中国人民大学人口学系讲师巫锡炜；中国台湾"中央"研究院社会学所助理研究员林宗弘；南京师范大学心理学系副教授陈陈；美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛；美国加州大学洛杉矶分校社会学系博士研究生宋曦。

关于每一位译者的学术背景，书中相关部分都有简单的介绍。尽管每本书因本身内容和译者的行文风格有所差异，校对也未免挂一漏万，术语的标准译法方面还有很大的改进空间，但所有的参与者都做了最大的努力，在繁忙的学习和研究之余，在不到一年的时间内，完成了三十五本书、超过百万字的翻译任务。李骏、叶华、张卓妮、贺光烨、宋曦、於嘉、郑冰岛和林宗弘除了承担自己的翻译任务之外，还在初稿校对方面付出了大量的劳动。香港科技大学霍英东南沙研究院的工作人员曾东林，协助我通读了全稿，在此

我也致以诚挚的谢意。有些作者，如香港科技大学黄善国教授、美国约翰·霍普金斯大学郝令昕教授，也参与了审校工作。

我们希望本丛书的出版，能为建设国内社会科学定量研究的扎实学风作出一点贡献。

吴晓刚

于香港九龙清水湾

序

迈克尔·豪特(Michael Hout)撰写的《流动表分析》一书全面回顾了分析职业出身与职业获得的交互分类数据时使用最广泛的方法。但是,这不应该使读者认为本书只对社会学家有用,或者只对那些使用这些方法的社会学家有用。正如豪特教授在本书开头部分所指出的以及在本书行将结束时所论述的那样,流动表分析的模型和方法可以被应用于相当广泛的研究领域。流动表的显著特征在于行分类与列分类的一致性[①]。既然这一特征在现实社会中如此常见,那么,豪特对流动表分析方法的回顾和解析就应该能够应用于许多不同的背景和学科。

第 1 章介绍了流动表分析的基本概念和检验方法,以说明是否存在完全流动(事实情况是,这种情形很少发生)。第 2 章考察了部分流动的几种特定情形,也就是说,流动在最高或最低类别中是有限的。这些模型是更为一般化的模型的特例,而如何发展这一模型则在第 3 章中做了交待。需要指

① 而不在于分类本身。——译者注

出的是，第 2 章所介绍的模型都假定行/列分类确实存在高低排序，但并非所有的流动表分析都要求这种假定。第 4 章提到的模型更为一般化，所涉及的流动性和稳定性情形更加复杂。第 5 章的关注点稍有变化，更接近于社会经济地位的获得分析。其方法和假设都与流动研究有所不同，豪特在本章一开始就对此作出了清楚的说明。第 6 章对最近的文献进行了讨论，向读者呈现出流动表分析将来的发展方向。最后，第 7 章通过几个实例展示了如何将流动表分析方法应用于其他研究领域。

作者在整本书中都留意并指出了各种模型之间的异同。为了实现这一目的，书中的模型都被应用于父亲职业与儿子第一份职业之间的同一个流动表数据，该数据来自费瑟曼(Featherman)和豪斯(Hauser)在 1973 年有关美国劳动力代际流动的研究。考虑到许多读者可能会尝试去重复书中的分析以更好地理解这些模型，本书在给出模型结果时保留了足够的细节，以便有兴趣的读者可以实现他们的这一想法。尽管本书并没有就如何运用计算机软件程序来估计这些模型给出详细指导，但还是就运算过程中的疑难之处提供了相关的建议。

鉴于本书所介绍的方法在流动研究乃至该领域之外具有广泛的应用性，并且全书的介绍全面、清晰，《流动表分析》一书对广大学生和从事研究工作的读者来说将会有很大的帮助。

理查德·G. 尼米

目 录

第1章

流动表分析

社会流动是社会科学的主要研究对象之一。对代际流动的研究兴趣可以一直追溯至马克思(Marx)和韦伯(Weber),但相比社会学的其他分支,该研究领域的分析技术却发展得更快,新发现和新成果也更多。这在一定程度上是由流动分析在增进我们对于社会分层的理解方面的地位所决定的。由于代际流动的结构性特征反映了资源占有的优劣势对比在上下两代人之间的持续,因而对于机会、阶级和特权等基本问题的回答都依赖于对这一流动结构的正确刻画。要领会这一点,可以尝试比较一下李普塞特和本迪克斯(Lipset & Bendix, 1959:57—60)、伦斯基(Lenski, 1966:411—415)、布劳和邓肯(Blau & Duncan, 1967:432—441)、赖特和佩罗内(Wright & Perrone, 1977)、费瑟曼和豪斯(Featherman & Hauser, 1978: 481—495)、布里格(Brieger, 1981)以及豪特等人关于资本主义社会机会结构的不同结论。

在过去的10多年里,方法论上的许多发展促进了社会流动研究的进步。其中的某些发展是相当技术化的。在本书中,虽然我努力将技术层面的细节保持在最低限度,但读者至少要对列联表数据分析技术的最新发展有必要的认识,

这样才能理解后面将要介绍的一些流动模型。在这方面，费恩伯格(Fienberg, 1980)、诺克和伯克(1980)、克洛格(Clogg)的著述无疑是很好的入门读物。另一方面，读者不必掌握潜在结构模型、准统一关联模型或其他类似模型的最新进展。当涉及这些模型时，它们会被作为新的参考资料在本书中予以介绍。

流动表分析的发展迅速扩展至其他研究领域，并见证了路径分析的日渐流行。本书所介绍的这些模型可以应用于任意一个对称表。一些并非社会流动分析的应用举例会在第 7 章中给出。

第1节 | 基本概念

流动表依据人们在两个时点上所从事的职业将他们进行划分。较早的时间点通常被称为“职业出身”，较晚的时间点则被称为“职业获得”。职业的高低次序排列通常依据的是社会经济地位，但这种排列反映的是社会流动模式的某种特征，而非流动表本身的特征。最常见的职业出身变量是受访者在成长时期，其父亲所从事的职业，另一个常见的职业出身变量是受访者的第一份职业。最常见的职业获得变量则是受访者当前所从事的职业(或简称“现职”)，而受访者的第一份职业也可以被当做职业获得变量，正如本书中许多研究实例所显示的那样。从父代职业向子代第一份职业或现职的社会流动是代际流动，而从(子代)第一份职业向现职的社会流动则是代内流动。

在选择了职业出身变量和职业获得变量之后，就必须定义各种职业类别，并将其进行编码。这是一项繁重的数据精简工作，因为原始的职业编码与流动表的职业编码相比要精细、具体得多(Hodge, 1981)。在美国劳工部的《职业名称词典》(*Dictionary of Occupational Titles*)中，就编撰了超过12000种职业。如此庞杂的职业分类体系在数量上必须被精简至可以用来研究的几个大类。许多大型研究项

目都把最初的职业数据信息按照美国人口普查所使用的441种分类体系重新进行编码(Featherman et al.,1977),但即使如此,对流动表分析来说,这一数字还是太大了。

在选择一个合适的职业分类数目方面,标准并不是很明确。许多研究者都是从理论和实际两方面的考虑来做决定。如果研究目的是识别阶级之间的界限,正如社会流动模式所揭示的那样,那么,职业分类就不应该将不同阶级的人们混在一起。例如,研究者必须慎重决定到底将“工头”划归为哪一个职业类别。在人口普查的职业分类体系中,“工匠”这一类别包括了“工头”和“技术工人”这两种具体职业。虽然“工头”与“白领工人”相比,更多的是从事体力工作,并且在社会经济地位上更接近于大多数的“工匠”,但就管理与被管理的关系而言,他们显然处于不同的阶级位置。

另外一项需要考虑的工作就是对已有研究的职业编码规则进行比较。如果辨别发展趋势或重复以往的研究结果构成了一个重要的目标,这种比较研究就显得尤其重要。在这种情况下,编码时最主要的考虑就应该是尽量与其他研究保持一致(Baron, 1980)。

本书的大部分例子都会将职业划分为五大类别。另外一些研究会使用17种分类法。关于五种分类法和17种分类法之间的对应关系,参见图1.1。但是,在英国社会流动分析中所使用的职业分类与此不太相同,因而也很难将其整合起来,这个例子反映出职业分类编码这项起步工作是何等重要(Goldthrope & Hope, 1972)。

上层非体力职业	专业人士(自雇)
	专业人士(受雇)
	经理人员
	销售人员(非零售业)
下层非体力职业	小业主
	文职人员
	销售人员(零售业)
上层体力职业	技术工人(制造业)
	技术工人(其他行业)
	技术工人(建筑业)
下层体力职业	服务业人员
	操作工人(其他行业)
	操作工人(制造业)
	搬运工人(制造业)
	搬运工人(其他行业)
农民	农场主或农场经营者
	农业耕作者

图 1.1 职业的五种分类法与 17 种分类法

一旦划分好职业类别，对数据的处理分析就可以开始了。在大多数研究中，同时也在贯穿本书的所有实例中，对职业出身和职业获得会采取同样的分类方法，而且排列次序也是相同的。流动表中的每一个单元格都对应一种职业出身和一种职业获得的组合，其数字就是具备这种组合的人数。如果用 i 表示行，用 j 表示列，那么 f_{ij} 就表示职业出身是 i 并且职业获得是 j 的人数[①]。当 i 和 j 相同时，这部分人的职业出身和职业获得就相同(前提是所使用的职业分类标准能够区分职业出身和职业获得)，这些 i = j 的单元格就构成了流动表的主对角线。这些单元格有特别的含义，因为它们代表的是不流动的一群人。对本书后面要介绍的许多社会流动模型而言，区分“流动的”和“不流动的”个体十分重要。

① 换言之，就是流动表中第 i 行、第 j 列的频数。——译者注

有时候，流动表也会出现某种不同的形式。与报告职业出身和职业获得不同，一些研究者会将受访者划分为向上流动、不流动、向下流动几个类别，然后将根据这种划分所构建的“流动状态”变量与职业出身、职业获得或一些其他变量进行交互分析。但这种操作方法不值得提倡，因为流动状态经常掩盖了数据中所存在的重要关系，它在简单合并职业出身和职业获得两方面信息的同时，忽略了这两个变量之间的相关性（Duncan，1966；Boudon，1975）。例如，采用上述这种分析路径的大多数研究都报告说，社会流动与教育和父代职业无关。构建流动状态这一变量的做法实际上仅仅采用了流动这一单方面的信息来报告数据，从而规避了职业出身和职业获得的对应和组合模式这一更为重要的研究问题。使用流动状态这种处理方式的问题就在于，它简单粗暴地置原始数据所存在的规律性于不顾。

表1.1给出了一个流动表的例子。该表根据父亲的职业类别和儿子的第一份职业类别（仅限全职性的平民工作），将1973年美国男性中20岁至64岁的就业人口进行了划分。本书中所举的例子，大部分都使用的是这个流动表。

表1.1　父亲的职业与儿子的第一份职业的交互分类

	儿子的第一份职业					
父亲的职业	上层非体力	下层非体力	上层体力	下层体力	农民	合计
上层非体力	1414	521	302	643	40	2920
下层非体力	724	524	254	703	48	2253
上层体力	798	648	856	1676	108	4086
下层体力	756	914	771	3325	237	6003
农民	409	357	441	1611	1832	4650
合计	4101	2964	2624	7958	2265	19912

资料来源：Featherman & Hauser，1978：49。

第2节 | 流动表数据的来源

大多数流动表数据都来自对劳动人口的调查。美国人口普查局于1962年和1973年收集了社会流动的数据，作为对3月“人口现状调查”数据的补充。两次数据收集的工作分别是在彼得·M. 布劳(Peter M. Blau)和奥迪斯·邓肯(Otis Duncan)、戴维·L. 费瑟曼(David L. Featherman)和罗伯特·M. 豪斯(Robert M. Hauser)的指导下进行的。10年一次的人口普查数据并没有提供代际流动的信息，但却能用于代内流动的分析，因为自1970年以来，人口普查便开始在样本人群中同时收集他们当前的职业和前一个职业的信息。

流动表数据的另一个来源是美国一些州的行政数据库中的婚姻记录。在这方面，罗格夫(Rogoff, 1953)利用印第安纳波利斯市的婚姻登记数据所做的代际流动分析便是一项具有开拓性的工作。

由于学者们所做的社会流动研究实在太多，在此恕难一一列出。值得注意的两项早期研究分别是格拉斯(Glass, 1954)在英国以及李普塞特和本迪克斯(Lipset & Bendix, 1952)在美国奥克兰进行的，其时间恰好都是在1949年。当代大规模收集社会流动数据的努力之一是美国综合社会调查(NORC, 1980)，该项目定期调查代际流动方面的信息。

第3节 | 初步分析

有一套适用于任何列联表的基本方法可以被用来对流动表进行初步分析。其中，最基础的分析方法就是计算行内或列内的百分比分布。表1.2报告的就是表1.1的行内百分比和列内百分比的计算结果。我们通常将行内百分比称为“流出”百分比，将列内百分比称为“流入”百分比。这两个百分比

表1.2 父亲的职业与儿子的第一份职业之间的流动:流入和流出百分比

	儿子的第一份职业					
父亲的职业	上层非体力	下层非体力	上层体力	下层体力	农民	合计
流入百分比						
上层非体力	34.5	17.6	11.5	8.1	1.8	14.7
下层非体力	17.7	17.7	9.7	8.8	2.1	11.3
上层体力	19.5	21.9	32.6	21.1	4.8	20.5
下层体力	18.4	30.8	29.4	41.8	10.5	30.1
农民	10.0	12.0	16.8	20.2	80.9	23.4
合计	100.0	100.0	100.0	100.0	100.0	100.0
流出百分比						
上层非体力	48.4	17.8	10.3	22.0	1.4	100.0
下层非体力	32.1	23.3	11.3	31.2	2.1	100.0
上层体力	19.5	15.9	20.9	41.0	2.6	100.0
下层体力	12.6	15.2	12.8	55.4	3.9	100.0
农民	8.8	7.7	9.5	34.6	39.4	100.0
合计	20.6	14.9	13.2	40.0	11.4	100.0

共同反映了代际之间的劳动力流动情况。流出百分比记录的是职业出身相同的人分别获得哪些不同的职业类别，反映的是劳动力流出特定职业类别的情况。流入百分比记录的是职业获得相同的人分别来自哪些不同的职业类别，反映的是劳动力流入特定职业类别的情况。数据显示出，在美国，不同的职业类别在流入和流出两方面都具有相当大的差异。即使是对特定的职业类别来说，流入和流出情况也大有不同。

仅仅计算流入和流出百分比只是一种非常粗浅的分析手段。虽然它们所提供的信息有助于理解社会分层的过程，但不同的社会阶层在劳动力供给和需求上的相对数量消长以及它所连带的其他社会过程也会影响最终的流入和流出模式。例如，农民的高生育率以及美国劳动力中农业人口比例的下降这两个原因，最终导致了农民这一职业类别在流入/流出模式上的失衡。虽然只有 39.4%的农民后代与他们的父亲一样继续从事农业工作，但 1973 年时，却仍然有 80.9%的农业工作者出身于农民家庭。本书所介绍的统计建模技术要实现的目标之一，就是控制美国（或其他国家）经济结构变化所带来的劳动力需求重新配置以及不同社会群体在生育率上的差异等外部力量的影响。

另一个初步分析方法是计算相异指数，它既可以指两类职业出身之间流出比例的相异性，也可以指两类职业获得之间流入比例的相异性，具体计算方法是把每一组对应比例之间的顺差加总。例如，下层体力职业类别与农民职业类别之间流出比例的相异指数的计算方法为：

$$\Delta = (0.126 - 0.088) + (0.152 - 0.077) +$$

$$(0.128-0.095)+(0.554-0.346)$$
$$=0.354$$[1]

这个相异指数测量的就是为使这两类职业的流出比例分布达到相同所需重新划分的人数比例。表 1.3 给出了每两类职业出身之间和每两类职业获得之间的相异指数。

表 1.3 父亲的职业与儿子的第一份职业之间的流动：每两类职业出身之间(对角线下方)及每两类职业获得之间(对角线上方)的相异指数

	职业获得类别				
职业出身类别	上层非体力	下层非体力	上层体力	下层体力	农民
上层非体力	—	0.169	0.310	0.353	0.483
下层非体力	0.163	—	0.155	0.192	0.688
上层体力	0.308	0.200	—	0.158	0.640
下层体力	0.384	0.276	0.157	—	0.607
农民	0.499	0.407	0.367	0.354	—

① 等式中的括号在原文中没有，为译者所加。—— 译者注

第4节 | 流动表的卡方检验

还有一个常用的初步分析工具就是卡方检验。在列联表数据分析中，卡方检验的零假设一般都是不存在相关关系的，这一点当然也适用于流动表。事实上，卡方检验所假设的统计独立模型在流动表分析这一研究背景下更具实质性的意义。

假定流动表中的职业出身变量分布在考虑了父代的劳动力需求状况和各职业群体的出生率之后是不变的，再进一步假定流动表中的职业获得变量分布在考虑了某国经济结构对子代的劳动力需求特征之后也是不变的。如果子代仅仅是在职业出身和职业获得这两个边缘分布的限制下随机进入到不同的职业中去，也就是说，当职业获得与职业出身这两个变量在统计上相互独立时，就实现了完全社会流动，因为个体流动到任意一类职业位置上的发生比完全是由职业获得变量的边缘分布所决定的。这就是为何大家所熟悉的统计独立模型经常被称为"完全流动模型"的原因。在完全流动模型中，每一类职业出身(即每一行)的流出百分比是完全相同的，因此，每两类职业出身之间的相异指数全都为0(对职业获得这一列变量而言也是如此)。当然，在现实的流动表中，每一行或每一列的比例分布很难完全相同。因此，

研究者就要借助卡方检验来判定实际观察到的流动模式与完全流动模型之间的差异是否可归结为仅仅由抽样误差所导致。

对完全流动模型进行卡方检验的第一步是计算该模型成立时的期望频数。具体计算方法与一般的卡方检验并无二致：

$$F_{ij} = n_i n_j / N \qquad [1.1]$$

上式中的N表示全部样本数，n_i和n_j分别表示职业出身变量和职业获得变量的边际频数①。流动表卡方检验的唯一特殊之处就是独立模型（即不存在统计相关）具有完全流动模型的意义。表1.4给出了根据美国流动表数据计算得到的期望频数。

表1.4 父亲的职业与儿子的第一份职业之间的流动：完全流动模型下的期望频数

	儿子的第一份职业					
父亲的职业	上层非体力	下层非体力	上层体力	下层体力	农民	合计
上层非体力	601.39	434.66	384.80	1167.00	332.15	2920
下层非体力	464.02	335.37	296.90	900.43	256.28	2253
上层体力	841.54	608.22	538.45	1633.00	464.78	4086
下层体力	1236.36	893.58	791.07	2399.15	682.84	6003
农民	957.70	692.18	612.78	1858.41	528.94	4650
合计	4101	2964	2624	7958	2265	19912

常见卡方检验的第二步是根据下面的公式将期望频数和观测频数进行比较：

$$X^2 = \sum_i \sum_j (f_{ij} - F_{ij})^2 / F_{ij} \qquad [1.2]$$

---

① 即第i行样本数的加总和第j列样本数的加总。——译者注

在大样本且零假设为真的条件下，该统计量接近于卡方分布，自由度为$(R-1)^2$(R等于列联表中的行数或列数)。在许多情况下，人们都将公式1.2直接视为“卡方”本身，但这是不准确的。尽管当零假设为真且样本量足够大时，该统计量确实接近于卡方分布，但卡方分布的计算公式本身并不是这样的。因此在本书中，我将把公式1.2中的统计量称为X^2。另一个在同等条件下也接近于卡方分布的统计量也经常被用到：

$$L^2 = 2\sum_i \sum_j f_{ij} \log(f_{ij}/F_{ij}) \qquad [1.3]$$

上式中的log表示的是取自然对数。L^2与X^2在计算公式上大相径庭，但它在零假设为真且样本量足够大的条件下也满足自由度为$(R-1)^2$的卡方分布。

如果关于完全流动的零假设是错的，那么，X^2与L^2两个统计量的计算结果都将大于相应的卡方值。因此，检验完全流动模型的方法就是计算X^2和L^2，并将它们与根据卡方分布表所查到的相应卡方值进行比较。如果X^2和L^2都小于相应的卡方值，我们就无法拒绝关于完全流动的零假设；如果X^2和L^2都大于相应的卡方值，我们就可以拒绝关于完全流动的零假设；如果相应的卡方值恰好在X^2与L^2之间(这种情况很少见)，检验的结果就不确定。

引进L^2统计量的目的并非是要核对使用X^2作为统计量的结果，也并非是为应对可能出现的卡方值恰好处于X^2与L^2之间的情形，而在于它在统计上是可以被分解的，且这种统计分解具有实质性的意义(但X^2却并不具备这种特性)。其实，L^2的确是更好的统计量，X^2之所以也被用到，是因为

它为人们所广泛熟知。将上述计算方法应用于表 1.1 的数据后，得出当自由度等于 16 时，X^2 值为 7166.77，L^2 值为 6170.13。再查表得知，在 16 个自由度下，卡方分布的 95% 分位数上的相应值是 26.30。因此，关于完全流动的零假设明显被拒绝。

另一个反映模型拟合度的指标是实际频数和期望频数之间的相异指数。虽然这一相异指数并没有检验关于完全流动的零假设(或其他任何零假设)，但由于它可以被解释为被模型错误划分的人数比例，因此对研究者来说也相当有吸引力。根据表 1.1 所报告的实际频数和表 1.4 所报告的期望频数，它们之间的相异指数为 0.201。这表示，完全流动模型对表 1.1 中的个案划分出现了约 1/5 的错误。

第 5 节 | 优比

一个较为深入的分析是计算相邻行与列的优比，它会在本书随后的章节中显示出其重要性。对一个 2 × 2 的列联表而言，表中的相关关系可以用优比或叉积比来测量：

$$\alpha = f_{11} f_{22} / f_{12} f_{21} \qquad [1.4]$$

有时也会使用 α 的变式（例如，取 α 的自然对数）（Goodman，1969a、1969b、1979a；Davis，1974；Fienberg，1980）。若 α 大于 1，则表示正相关；若 α 小于 1，则表示负相关。由于任何一个流动表都是由许多这种相邻的行与列所组成的 2 × 2 表所构成，所以一个 R × R 表中的相关关系就可以用 $(R-1)^2$ 个优比来表达（Goodman，1979a）。值得注意的是，行与列的分类次序会直接影响优比的计算结果。古德曼（1979a）将这些 2×2 表称为列联表的"基本单元"。除此之外，所有其他的分表所包括的信息都是多余的。因此，通过如下公式计算得出 $(R-1)^2$ 个优比或者它们的变式值，就获得了流动表中所有的相关关系信息。

$$\alpha_{ij} = f_{ij} f_{i+1,\, j+1} / f_{i+1,\, j} f_{i,\, j+1} \qquad [1.5]$$

请注意，在检验完全流动模型时有多少自由度，就有多少 2 × 2 表和多少优比。这并不是巧合。针对同一个流动表

中所蕴含的相关关系的任意一个统计模型设计，都可以由该模型所预示的重要优比的数目来概括(如果 $\alpha = 1.0$，就说明它是一个不重要的优比)。完全流动模型意味着所有的 α_{ij} 都等于 1.0。在随后的章节中我们会看到，完全流动模型以及比它更复杂的流动模型下的优比所呈现出来的不同模式构成了模型之间相互比较的基础。正如前文所言，优比的模式取决于职业类别的排列次序。并不是所有的排列次序都同样适用于统计建模，但只要研究者采用了“正确的”排列次序，优比就是有用的。从美国流动表数据所观察和计算得到的优比都显示在表 1.5 中。值得注意的是，优比的四个最大取值都出现在对角线上。要设计更为精致复杂的理论模型，必须充分考虑到这一重要信息。

表 1.5　表 1.1 中基本 2×2 分表的优比值

	职业获得比较			
职业出身比较	5∶4	4∶3	3∶2	2∶1
5∶4	1.964	0.836	1.300	1.098
4∶3	1.122	2.725	0.707	0.944
3∶2	1.489	0.639	2.203	1.106
2∶1	0.722	1.464	0.847	15.954

第6节 | 流动比率

在介绍比完全流动模型更复杂的统计模型之前，我想先谈谈流动比率。流动比率这一概念由戈尔达梅(Goldhamer)首先提出，并经由罗格夫的推广而流行。对流动表中的任意一个单元格来说，流动比率指的就是这个单元格的实际频数与完全流动模型假设下的期望频数之比：

$$R_{ij} = f_{ij}/F_{ij} \qquad [1.6]$$

尽管 R_{ij} 在进行探索性分析时确实有用(例如，Hauser，1979)，但在更多的情况下，它却是被当做测量职业出身和职业获得两个变量之间相关关系的一个指标来使用。遗憾的是，R_{ij} 其实并不能测量这种相关性。

将流动比率视为相关性的测量指标这种做法，表面上看似乎顺理成章：既然它表示的是实际频数与不存在相关关系情况下的期望频数之间的比值，那么，它作为一个指标当然就能反映出相关程度的大小。然而，仔细分析一下就会发现，这种理解其实站不住脚，具体而言有三个原因。

第一，正如布劳和邓肯(1967：93—97)所展示以及泰里(Tyree，1973)所阐释的那样，如果两个流动表所蕴含的职业出身和职业获得之间的相关模式完全相同而仅仅在边缘分

布上有所差异，那么它们的流动比率则必然不同。这也是流动比率概念遭受批评最多的一点。对流动比率概念的另外两个批评意见都或多或少与此有关。

第二个批评指出，流动比率是基于完全流动模型计算出来的，而这一模型却并不拟合大多数的经验数据(Goodman，1965、1969b；Hauser，1978、1979、1981)。既然它的模型基础是错的，那么，它所反映的信息就同时囊括了系统性的相关和误差。这是因为，每一个单元格的频数都由边缘效应、系统性相关和误差(包括测量误差和抽样误差)共同决定。流动比率概念的拥趸认为，它反映了系统性相关和边缘效应的相对情况，但却并不承认误差的存在及其重要性。

尽管计算流动比率时会出现误差，但如果误差相对于系统性相关而言并不严重，并且按照完全流动模型所计算的边缘效应是对真实边缘效应的无偏估计，那么，流动比率作为职业出身和职业获得之间的相关性测量指标就仍然是可以接受的。然而，上述条件在真实数据中却很少成立。

对流动比率的第三个批评是，实践中如果没有一个与数据相拟合的模型用来作为参照，研究者就无从评估误差的相对重要性和系统性的相关到底有多大。而一旦找到了能够拟合数据的这一模型，就不再需要计算流动比率了，因为该模型的参数本身已经可以更好地反映流动表所蕴含的相关性。

霍普(Hope，1981)曾试图为流动比率概念做辩护，其理由是，对任何给定的流动表而言，它的取值都是不变的，而其他替代性参数的取值则都随模型的拟合度而不同。但流动比率的问题恰恰就在于它的不变性：不管与给定的流动表相

拟合的正确模型到底如何，流动比率永远保持同样的取值。因此，流动比率最多只能提供一个并不完整的线索，以探索流动表中所隐藏的相关关系结构（Hauser, 1979）。更糟糕的是，这一概念经常会起误导作用。总而言之，流动比率不应该被当做对相关性的测量。

第2章

流动表的部分独立

第 1 章有关卡方检验的部分已经提到，完全流动模型并不能拟合表 1.1 中的美国社会流动数据。我现在就介绍更为复杂的模型。本章将要涉及的模型将完全流动仅限定于职业出身和职业获得的某些组合，而将其他组合情况视为存在流动障碍。在表格的完全流动区域中，职业获得与职业出身具备统计上的独立性，但在表格的其他区域，它们之间却存在显著相关。许多最近的社会流动研究文献都探讨了这一议题，即存在于某些职业类别之间的完全流动以及横亘于其他职业类别之间的流动障碍（例如，Hauser，1978、1979；Breiger，1981）。

寻找流动障碍的首要着眼点在流动表的对角线位置。美国流动表数据中实际频数和期望频数之差表明，真实社会中所存在的不流动情形要远甚于完全流动模型的假设。这一特征确实很常见。社会流动研究已经不断证实了这种“过度的”不流动性的存在（例如，Lipset & Bendix，1959；Blau & Duncan，1967；Featherman & Hauser，1978）。古德曼的准完全流动模型（QPM）（1961、1965、1968、1969a、1969b、1972a）针对的正是这一不流动情形，它会被首先加以介绍。我也会提及准完全流动模型的一般化形式，它加入了对非对

角线位置上的向上或向下流动类型所作出的约束。这又具体包括两个模型，其一是允许流动障碍还出现在职业地位序列的两端(Goodmand，1965、1972a；Pullum，1975)，其二是允许存在系统性的流动障碍。

第1节 | 定义准完全流动

为分析对角线上集中了过多个案的流动表，几位研究者(Blumen et al.，1955；Goodman，1961、1965、1969a、1969b；White，1963)都提出了将流动和不流动加以区分的流动模型。以古德曼的模型为例，观察到的流动模式被视为两个过程的共同结果。第一个过程是不流动部分，它使得特定的一群人——"滞留者"——处于与其出身相同的职业位置上。第二个过程是流动部分，类似于完全流动的情况，因为其余人群的职业获得不因职业出身而有所不同。在后一个过程中，某些个体仍然有可能从事与他们的出身相同的职业。虽然这些人貌似那些滞留者，但却与他们不同。"滞留者"之所以没有离开他们的职业出身，乃是由模型未能解释的某些结构性力量所致，而移动者中职业获得与职业出身相同的情况则纯属偶然，就像那些职业出身与职业获得不同的人一样，他们最终进入哪种职业类别都是随机的。

古德曼的模型其实是对布吕芒(Blumen et al.，1955)所提出的移动者—滞留者模型的进一步发展。最初的移动者—滞留者模型依据职业获得的不同对人群进行划分。凡是职业获得与职业出身相同的人都被划分为滞留者。根据这个模型，只有职业出身和职业获得不同的人才是被随机过

程所主导的,完全流动被认为是以移动事实的发生为条件的。虽然古德曼的准完全流动模型沿袭了移动者—滞留者模型及其扩展模型(Singer & Spilerman, 1974、1976)的一贯思路,但它的内在逻辑却更为严谨,同时也与其他流动模型具有更高的一致性。在移动者—滞留者模型中,只有滞留者才是不流动的。但在准完全流动模型中,移动者中的某些人也是不流动的,这是出自偶然因素的作用。他们是服从完全流动模式,但只不过恰好发生职业获得与职业出身相同这种情形的一群人。如若不然,在这群人中将会不可避免地出现职业出身与职业获得相关的情形,而这是完全流动所不允许的(按照定义,只有当职业出身与职业获得相互独立时,完全流动才得以存在)。

第 2 节 | 准完全流动的对数线性模型

在用文字说明了模型的基本原理之后，现在我将用更严谨的数学公式来进行表述。为此，先要简单介绍一下对数线性模型的数学表达方法。对一个两维列联表(包括流动表)而言，对数线性模型的一般形式为：

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + b_{ij} \qquad [2.1]$$

上式中的 a_0 表示总均值，a_{1i} 表示行效应，a_{2j} 表示列效应，b_{ij} 表示行与列的交互效应。这些参数服从下列约束：

$$\sum_i a_{1i} = \sum_j a_{2j} = 0 \qquad [2.2]$$

和

$$\sum_i b_{ij} = \sum_j b_{ij} = 0 \qquad [2.3]$$

当交互效应不存在时(即所有的 b_{ij} 都等于 0 时)，我们就得到了完全流动模型。如果所有的 b_{ij} 都不等于 0，我们得到的就是饱和模型，该模型用完了流动表数据中所有的自由度，并且与经验数据完全吻合。本书此后的章节所涉及的模型全都处于完全流动模型与饱和模型之间。这些模型都是对 b_{ij} 作出某种约束以避免它们都等于 0 或都不等于 0，也就是试图在并不用完所有自由度的情况下，获得一个可以接受的模

型拟合度。

公式 2.1 也可以被写为相乘形式：

$$F_{ij} = A_0 A_{1i} A_{2j} B_{ij} \quad [2.4]$$

与公式 2.1 一样，公式 2.4 中的参数也必须服从类似的两个约束，其相乘形式如下：

$$\prod_i A_{1i} = \prod_j A_{2j} = 1 \quad [2.5]$$

和

$$\prod_i B_{ij} = \prod_j B_{ij} = 1 \quad [2.6]$$

对数线性模型和相乘模型中的参数具有简洁的对应关系，即 $a_{1i} = \log(A_{1i})$，$a_{2j} = \log(A_{2j})$，$b_{ij} = \log(B_{ij})$。因此，当所有的 B_{ij} 都等于 1 时，我们就得到了完全流动模型，公式 1.1 与公式 2.4 在这种情况下就是等同的。为节省篇幅，本书将只采用对数线性的数学形式来对模型进行表述。

之所以需要公式 2.3 这一约束条件，是因为公式 2.1 中的参数个数超过了自由度的个数。这一约束条件的存在使我们不必估计 2R－1 个交互项参数，因而才能够对饱和模型进行估计。但对准完全流动模型来说，只有当对角线上的 b_{ij} 不等于 0 时，也就是说，不等于 0 的参数个数少于自由度的个数时，公式 2.3 这一约束条件就不需要了。准完全流动模型的对数线性方程表达如下：

$$\begin{aligned} \log(F_{ij}) &= a_0 + a_{1i} + a_{2j} + b_{ii}\text{①} \quad (\text{当 } i = j \text{ 时}) \\ \log(F_{ij}) &= a_0 + a_{1i} + a_{2j} \quad (\text{当 } i \neq j \text{ 时}) \end{aligned} \quad [2.7]$$

---

① 原文此处为 b_i，为笔误，已改正。——译者注

公式 2.2 这一约束条件仍然适用。公式中的 b_{ii} 表示的正是继承效应，它们是模型中仅有的交互项参数，测量了对角线单元格上的个案分布偏离完全流动模式的程度。适用于非对角线单元格（即 $i \neq j$ 的单元格）的公式与完全流动模型的公式是相同的，除了它必须要求 i 不等于 j 这一条件之外。准完全流动模型比完全流动模型多用了 R 个自由度——每增加一个 b_{ii} 就多用一个自由度，因此，它的自由度计算公式为：

$$df_{QPM} = (R-1)^2 - R$$

对准完全流动模型下 F_{ij} 的计算是通过"屏蔽"或删除对角线上的个案之后，再对剩余的流动表数据按独立模型进行估计得到的。然而，正是由于对角线上的单元格被排除在外，所以就不能再用公式 1.1 来计算剩余单元格的 F_{ij}。除了极特殊的情况以外（Bishop et al.，1975：192—206），一般都需要用到迭代方法。

第3节 | 准完全流动模型的结果

准完全流动模型也并不能很好地拟合表1.1中的美国流动表数据。无论以哪个常用的显著性水平来衡量，L^2和X^2都超过了相应的卡方值。但这并不意味着该模型所考虑的不流动情形不重要。因为与完全流动模型相比，准完全流动模型的这两个统计量都小得多，L^2就减少了5486.79。还记得完全流动模型错误地划分了20.1%的个案吗（见表1.1）？相比之下，准完全流动模型仅仅错误地划分了5.5%的个案。

对F_{ij}的估算见表2.1的上半部分。请注意，对角线上的数字并不是F_{ij}，而是由偶然因素所导致的不流动的个案数（F_{ij}^*）。这些估算结果是采用下述方法得到的。在前面的章节中，我已经介绍了优比这个概念，并且说明了在完全流动模型下，所有基本单元的优比均等于1.0的这种特性。在准完全流动模型下，完全流动只适用于移动者。因此，对角线上某个单元格的移动者个案数的估算原理就在于，凡是涉及这个单元格的所有基本单元的优比都必须等于1.0。让我们从流动表的中间位置开始。涉及对角线上第三个单元格的，有四个基本单元，它们的优比必须都等于1.0：

$$\alpha_{22} = F^*_{22}F^*_{33}/F_{23}F_{32} = 1.0$$
$$\alpha_{23} = F_{23}F_{34}/F_{24}F^*_{33} = 1.0 \qquad [2.8]$$
$$\alpha_{32} = F_{32}F_{43}/F^*_{33}F_{42} = 1.0$$
$$\alpha_{33} = F^*_{33}F^*_{44}/F_{34}F_{43} = 1.0$$

首尾两个方程都包含两个未知数,但中间的两个方程都只包含一个未知数:F^*_{33}。用中间两个方程中的任意一个就可以求出 F^*_{33}。例如,将第二个方程求解,即得到:

$$F^*_{33} = F_{23}F_{34}/F_{24} = (321.81)(1645.78)/(914.55) = 579.11$$

将这个数值分别代入首尾两个方程中,即可求得 F^*_{22} 和 F^*_{44}。以此类推,即可得出 F^*_{11} 和 F^*_{55},其计算公式分别为:

表 2.1 父亲的职业与儿子的第一份职业之间的流动:准完全流动模型的估计结果

	职业获得				
职业出身	上层非体力	下层非体力	上层体力	下层体力	农民
(1) 期望频数					
上层非体力	372.17	344.01	285.54	811.49	64.96
下层非体力	419.43	387.70	321.81	914.55	73.21
上层体力	754.79	697.68	579.10	1645.78	131.74
下层体力	934.37	863.67	716.88	2037.33	163.09
农民	578.41	534.64	443.77	1261.18	100.96
(2) 参数估计值					
上层非体力	1.335				
下层非体力		0.301			
上层体力			0.391		
下层体力				0.490	
农民					2.898

注:$L^2 = 683.34$; $X^2 = 720.74$; $df = 11$; $p < 0.01$; $\Delta = 0.055$。除非特别说明,本书表格中所报告的参数估计值都是就对数形式的方程而言。

$$F_{11}^* = F_{21}F_{12}/F_{22}^*$$

和

$$F_{55}^* = F_{45}F_{54}/F_{44}^*$$

由于本书在报告数据时，只保留了有限的几位小数，但上述计算过程却牵涉到小数位数的省略问题，因此，读者计算出来的结果可能会有少许出入。

对参数 b_{ii}（即相加形式方程中的继承效应）的估计结果见表 2.1 的下半部分，其计算公式如下：

$$b_{ii} = \log(f_{ii}/F_{ii}^*) \qquad [2.9]$$

一些统计软件，例如 FREQ（Haberman，1979），能够直接对 b_{ii}进行估算，省去了研究者手工计算的麻烦。表 2.1 给出的这些 b_{ii}的参数值说明，继承性在农民职业阶层中表现得最强（b_{55} =2.898），其次是上层非体力职业阶层（b_{11} = 1.335①），接下来依次是下层体力职业阶层和上层体力职业阶层，最弱的是下层非体力职业阶层。其实不用太在意这种次序，因为这一模型与数据并不拟合。况且，b_{22}和 b_{33}其实非常接近，而 b_{44}也与这两个数值相差不大。

① 原文为 b_{ii}，为笔误，已改正。——译者注

第 4 节 | 存在于顶端和底端的流动障碍：边角模型

准完全流动模型更一般化的形式是在屏蔽对角线上的单元格之后再屏蔽非对角线上的某些单元格。古德曼(1965)和普勒姆(Pullum, 1975)认为，准完全流动模型之所以并不拟合数据，是因为职业地位层级的两端存在“过多的”短距离流动。至于何为“短距离”和“职业地位层级的两端”，取决于流动表中职业分类的具体排列次序。他们假设，发生于下层非体力职业阶层与上层非体力职业阶层之间的流动，其实与准完全流动模型所预期的这两个职业阶层的继承性并没有太大的不同。他们还假设，与准完全流动模型所预期的情形相比，农民职业阶层的后代进入下层体力职业阶层的期望频数以及相反情形(即下层体力职业阶层的后代进入农民职业阶层)下的期望频数其实会更高。由于这种理论模型并没有一个公认的命名，我将称之为“边角模型”。

对边角模型的正式表述是，它不仅拟合了对角线上的 R 个参数(在这一点上与准完全流动模型相同)，并且还拟合了边角单元格(1, 2)，(2, 1)，(R－1, R)，(R, R－1)上的四个参数：

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + b_{ij}Z_{ij} \qquad [2.10]$$

上式中的 a_{1i} 与 a_{2j} 之和必须为 0（其道理与公式 2.2 相同），并且当 $i = j$ 或者第（i, j）格是一个边角单元格时，$Z_{ij} = 1$。与公式 2.1 相比，多出的参数 Z_{ij} 旨在令边角单元格上的实际频数与期望频数相等。既然每当 $Z_{ij} = 1$ 时就损失一个自由度[①]，它的自由度就是 $(R-1)^2-(R+4)$。

尽管对美国的流动表数据而言，边角模型的出发点是为了拟合过多的短距离流动，但其实该模型也可以拟合与此相反的情形，即过少的短距离流动。第（i, j）单元格上的个案数到底是过多还是过少，从 b_{ij} 的符号上可一见分晓：如果 b_{ij} 是正值，就表示第（i, j）格上的实际个案数比准完全流动模型所预期的要多；如果 b_{ij} 是负值，就表示观测频数比期望频数要少。

根据表 2.2 中的结果，边角模型比准完全流动模型更好地拟合了美国的流动表数据，但是，它的拟合度在常规的显著性水平下仍然尚未达到可以接受的水平。由于该模型要精确地拟合对角线上的单元格和边角单元格，所以，这些单元格上的个案数就是按照完全流动模型所估计的期望频数 F^*_{ij}。F^*_{ij} 的计算原理和过程与准完全流动模型类似。例如，计算时会用到方程 $F^*_{21} = F_{31}F^*_{22}/F_{32}$。得到 F^*_{ij} 后，模型中的每个参数值都可以用如下公式来计算：

$$b_{ij} = \log(f_{ii}/F^*_{ii})$$

在解释这些参数时务必小心谨慎，因为该模型与数据并

---

① 那么边角模型与准完全流动模型相比，就多用了 4 个自由度。——译者注

不拟合，当然它们也不是太离谱，这一点会在随后的分析中再加以说明。表2.2中的这些参数值表明，农民阶层的继承性非常之高，但上层非体力职业阶层的继承性在强度上降低了一半，而从下层非体力职业阶层向上层非体力职业阶层的流动性强度（与这种上层非体力职业阶层的继承性强度相比）又降低了一半。此外，有五个单元格上的参数值几乎相当，在数值大小上都处于中间位置，它们分别是第(1，2)格、第(4，5)格和第(5，4)格以及两个对角线上的单元格，即第(2，2)格和第(4，4)格。最后，边角模型中上层体力职业阶层的继承性与根据边缘效应所预期的情形大体一致（随后的分析会表明，b_{33}在统计上确实是不显著的）。

表2.2 父亲的职业与儿子的第一份职业之间的流动：边角模型的估计结果

	职业获得				
职业出身	上层非体力	下层非体力	上层体力	下层体力	农民
(1) 期望频数					
上层非体力	286.48	280.06	279.12	662.89	42.99
下层非体力	292.29	285.74	284.79	676.35	43.87
上层体力	727.22	710.92	708.56	1682.75	109.14
下层体力	826.90	808.37	805.69	1913.42	124.10
农民	408.88	399.71	398.39	946.13	61.36
(2) 参数估计值					
上层非体力	1.597	0.621			
下层非体力	0.907	0.606			
上层体力			0.189		
下层体力				0.553	0.647
农民				0.532	3.396

注：$L^2 = 50.12$；$X^2 = 50.44$；$df = 7$；$p < 0.01$；$\Delta = 0.014$。

第 5 节 | 对称相关

准完全流动模型在流动表分析中是一个重要的基准模型。许多研究者都以它为基础再加入更多的参数以使模型能够拟合数据。边角模型就体现了这种研究策略。另外一个重要的基准模型则是对每个单元格都拟合一个交互项参数,同时约束这些参数具有对称性的特征(Goodman，1968、1972c；Knoke & Burke，1980:49—54)：

$$F_{ij} = a_0 + a_{1i} + a_{2j} + b_{ij} \qquad [2.11]$$

上式也需服从公式 2.2 和公式 2.3 的条件约束,并且令 $b_{ii} = b_{ji}$。这就是大家所知道的“准对称模型”(QS)。应用这一模型,研究者就能用各个国家的流动表数据来检验职业阶层之间的流动是否具有对称性。如果准对称模型能够拟合数据,那么就不必再考虑非对称模型。如果准对称模型不能拟合数据,研究者可以渐次加入非对称参数,使最终模型的非对称参数的个数保持在最低限度。

有两个模型与准对称模型具有密切联系。第一个是对称模型(S),它不仅将交互项参数约束为对称,而且将边缘效应也约束为对称：

$$F_{ij} = a_0 + a_{1i} + a_{1j} + b_{ij} \qquad [2.12]$$

上式也需服从公式 2.2 和公式 2.3 的条件约束,并且令 $b_{ij} =$

b_{ji}。在对称模型下，边缘效应参数和边缘分布频数对职业出身和职业获得两个变量来说都相等。

第二个是边缘齐性模型(MH)。在这一模型下，交互效应并不对称，但职业出身与职业获得的边缘分布却相同。边缘分布相同与边缘效应相同是两码事。由于 MH 模型下的交互效应并不具有对称性，边缘效应就不可能相同，边缘分布本身也不可能相同。事实上，MH 模型并不是一个对数线性模型，但它可以用对数线性方法来检验，因为它的拟合度就等于 S 模型的 L^2 值和 QS 模型的 L^2 值之差。

表 2.3 显示了 S 模型和 QS 模型的期望频数以及 S 模

表 2.3 父亲的职业与儿子的第一份职业之间的流动：对称模型与准对称模型的期望频数

	职业获得				
职业出身	上层非体力	下层非体力	上层体力	下层体力	农民
(1) 对称模型					
上层非体力	1414	622.50	550.00	699.50	224.50
下层非体力	622.50	524	451.00	808.50	202.50
上层体力	550.00	451.00	856	1223.50	274.50
下层体力	699.50	808.50	1223.50	3325	924.00
农民	224.50	202.50	274.50	924.00	1832
(2) 准对称模型					
上层非体力	1414	575.88	289.06	599.01	42.06
下层非体力	669.12	524	264.16	752.30	43.42
上层体力	810.94	637.84	856	1657.81	123.40
下层体力	799.99	864.70	789.18	3325	224.13
农民	406.94	361.58	425.60	1623.87	1832
	L^2	X^2	df	p	Δ
对称模型	2804.87	2562.79	10	<0.01	0.118
准对称模型	27.45	27.35	6	<0.01	0.011
边缘齐性模型①	2777.42	—	4	<0.01	—

① 原文为“边缘对称模型”，但为了与正文部分的模型名称保持统一，这里译为“边缘齐性模型”，其实，两者所要表达的意思本质上是相同的。——译者注

型、QS 模型和 MH 模型各自的拟合度。然而，将 QS 模型和 QPM 模型进行比较就会发现，表 1.1 中非对角线单元格上的相关模式几乎都具有对称性。豪斯（1981）已经表明，只要将两个非体力职业阶层之间的（向上和向下）流动，即第（1，2）格和第（2，1）格设定为一组非对称参数，改进后的模型就能相当好地拟合数据（$L^2 = 7.79$；$X^2 = 7.78$；$df = 5$；$p = 0.26$；$\Delta = 0.005$）。这显示了 QS 模型作为一个诊断模型的作用（Hauser，1979）。邓肯（1981）还展示了在分析面板数据时，如何运用 S 模型、QS 模型和 MH 模型①。

① 原文为 HM，为笔误，已改正。——译者注



第3章

社会距离模型

前面所介绍的准完全流动模型和边角模型其实都可看做另一个更一般化模型的特殊形式，它是由古德曼(1972a)提出的。本章会介绍此一般化模型所涉及的更多的具体形式。这些模型都将职业类别按地位高低或其他维度进行排列。在大多数情况下，这种排列都假定相邻两个职业类别的间距是相同的。但最重要的假定是，某种社会流动的发生比是由它所跨越的职业类别界限所决定的。一些模型关心社会流动跨越了几个职业类别，另一些模型则关心社会流动跨越了哪些职业类别。不管是哪一种情况，如果职业类别的排列次序不清楚，则该模型就不适合。

第1节 | 对角线约束模型

本节介绍的模型是从古德曼(1972a:661—671)的跨对角线模型(DC)发展而来的:

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + c_{ij} + d_k + u_{ij} \qquad [3.1]$$

在上式中,

$k = i - j$

$c_{ij} = \sum v_s$ 当 $i > j$ 时,s 的取值为 j 到 $i-1$

当 $i < j$ 时,s 的取值为 i 到 $j-1$

$u_{ij} = \log(f_{ij} - a_0 - a_{1i} - a_{2j} - d_k)$ 当 $(i, j) \in S$ 时

$u_{ij} = 0$ 其他情况

并且适用公式 2.2。s 表示的是模型所要屏蔽的单元格。

最简单的对角线约束模型是将主对角线上的所有参数都设定为一个值(d_0),并用相同的职业类别分布来检验继承性是否真的超过了基于完全流动模式所预期的情形:

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_0 \qquad [3.2]$$

并且适用公式 2.2。这一模型被称为"准完全流动(约束)模型"(QPM-C)。这一模型只用了一个 d_k 值(公式 3.1 中本来应该有五个 d_k 值),并且没有用到 c_{ij} 或 u_{ij} 参数(即公式 3.1 中的这些参数被设定为 0)。与 QPM-C 模型相关的还有两个模型。

第一个模型是进一步对次对角线上的参数作出如下约束：

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_{|k|} \qquad [3.3]$$

在上式中，$|k| = |i-j| = 0$ 或 1，并且适用公式 2.2。这时，将次对角线上的所有参数也都设定为同一个值，因此该模型具有对称性，被称为“对角线对称（约束）模型”（SD-C）。第二个模型则是在第一个模型的基础上，放松次对角线上参数的这种对称性要求：

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_0 + d_k \text{①} \qquad [3.4]$$

上式中的 $k = i-j = -1$、0 或 1，并且适用公式 2.2。该模型被称为“对角线（约束）模型”（D-C）。

逐步放松对主对角线和次对角线参数的约束（即逐步加入 u_{ij} 项），就又产生了四个模型。其中的一对模型都是放松对主对角线参数的约束，即在公式中加入可能的五个 u_{ij} 项，分别见公式 3.5 和公式 3.6：

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_{|k|} + u_{ij} \qquad [3.5]$$

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_k + u_{ij} \qquad [3.6]$$

在上式中，当 $i = j$ 时，$(i, j) \in S$，并且适用公式 2.2。这两个模型分别用 QSD-C 和 QD-C 来表示，与前两个模型 SD-C 和 D-C 相比，多了一个字母 Q，它代表英文单词“quasi”，意思是屏蔽主对角线。另外一对模型是放松对次对角线参数的约束：

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_{|k|} + u_{ij} \qquad [3.7]$$

在上式中，$u_{ij} = u_{ji}$，此外，

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_k + u_{ij} \qquad [3.8]$$

① 原文此处的下脚标错误，已改正。—— 译者注

在上式中，当$-1 \leqslant i-j \leqslant 1$时，$(i, j) \in S$，并且适用公式2.2。这两个模型被简写为QSD和QD。QSD并不属于公式3.1所定义的模型系列，但它仍不失为一种理论概括。

上述七个模型之间的异同可参见图3.1。图中用一些整数来代表交互项参数或每个单元格上交互项参数的组合。如果某个单元格上的数字等于另外一个单元格上的数字，就表明这两个单元格上的交互项参数是相等的。古德曼

1				
	1			
		1		
			1	
				1

QPM-C

1	2			
2	1	2		
	2	1	2	
		2	1	2
			2	1

SD-C

1	2			
3	1	2		
	3	1	2	
		3	1	2
			3	1

D-C

1				
	2			
		3		
			4	
				5

QPM

1	6			
6	2	6		
	6	3	6	
		6	4	6
			6	5

QSD-C

1	6			
7	2	6		
	7	3	6	
		7	4	6
			7	5

QD-C

1	6			
6	2	7		
	7	3	8	
		8	4	9
			9	5

QSD

1	6			
7	2	8		
	9	3	10	
		11	4	12
			13	5

QD

图3.1 对角线约束模型的交互项参数设置

(1979b:808)提议将这种图示“作为不断启发思考的一个工具来使用,从而建构出更多的模型,以尽可能研究各种相关的、真切的社会现象”。

关于这七个模型的结果见表3.1和表3.2。表3.1虽然只显示了三个模型的期望频数,但给出了每个模型的拟合优度(完全流动模型和准完全流动模型的拟合优度也一并列出以做比较);表3.2报告了所有模型的参数估计值。正如表3.1所显示的那样,没有一个模型在0.05的显著性水平上能够拟合数据。但需要说明的是,OCG-II的样本规模增加了研究者寻找与数据相拟合的模型的难度。该调查的样本量实在太大,以至于对期望频数的任何微小的偏离在统计检验上都是显著的。样本量对统计显著性的这种影响的重要性从Δ值的变化上可见一斑。从QPM模型到PM模型,被错误地划分的个案比例仅仅减少了1/4(从20.1%下降到5.5%),但QSD模型进一步将个案错误地划分比例降低了1/4(从5.5%下降到1.3%),QD模型更是将它降低了一半(从1.3%下降到0.6%)。

从表3.1和表3.2可以得出的第一个结论是,职业的代际继承性是很强的,尤其是在职业序列的两端(即顶端的上层非体力职业阶层和底端的农民职业阶层)。u_{ij}出现负值并不意味着不存在职业地位的继承性(Goodman, 1965),因为它们表示的是对相应d_k值的偏离情况,而且,既然u_{ij}的和被约束为0,那么自然会有一些负值出现。d_0和u_{ij}[①]两个交互效应相加为负值的情形,只在QSD和QD两个模型的第三个对角线单元格上出现,但数值本身并不大。

---

① 原文此处为u_{ii},为笔误,已改正。——译者注

表 3.1 父亲的职业与儿子的第一份职业之间的流动：对角线约束模型的估计结果

	职业获得				
职业出身	上层非体力	下层非体力	上层体力	下层体力	农民
（1）QPM-C 模型的期望频数					
上层非体力	1305.68	334.77	193.18	831.09	255.28
下层非体力	372.39	814.64	160.94	692.37	212.67
上层体力	750.88	562.35	947.88	1396.07	428.82
下层体力	857.47	642.18	489.69	3523.97	489.69
农民	814.58	610.06	352.04	1514.50	1358.03
（2）SD-C 模型的期望频数					
上层非体力	1414.28	474.89	224.13	618.88	187.81
下层非体力	553.00	768.31	290.07	492.24	149.38
上层体力	615.02	683.51	1067.74	1449.41	270.32
下层体力	746.14	509.62	636.82	3576.80	533.62
农民	772.56	527.67	405.23	1820.67	1123.87
（3）QSD-C 模型的期望频数					
上层非体力	1414	481.92	243.66	718.48	61.95
下层非体力	590.48	524	361.34	715.49	61.69
上层体力	639.21	773.66	856	1717.68	99.45
下层体力	906.04	736.38	825.66	3325	209.92
农民	551.27	448.04	337.34	1481.35	1832
	L^2	X^2	df	p	Δ
PM	6170.13	7166.77	16	<0.01	0.201
QPM-C	2480.14	2266.81	15	<0.01	0.125
SD-C	1955.97	1905.68	14	<0.01	0.112
D-C	1952.96	1902.62	13	<0.01	0.113
QPM	683.34	720.74	11	<0.01	0.055
QSD-C	333.80	333.00	10	<0.01	0.042
QD-C	327.857	327.79	9	<0.01	0.041
QSD	33.63	33.52	7	<0.01	0.013
QD*	15.54	15.48	3	<0.01	0.006

表 3.2 父亲的职业与儿子的第一份职业之间的流动：对角线约束模型的交互项参数估计值

参数	QPM-C	SD-C	D-C	QPM	QSD-C	QD-C	QSD	QD
d_0	1.072	1.197	1.194	1.083	1.262	1.256	1.226	1.233
d_{iu}	—	0.487	0.456	—	0.398	0.350	0.233	0.219
d_{id}	—	0.487	0.505	—	0.398	0.422	0.233	0.254
u_{11}	—	—	—	0.252	0.005	0.005	0.483	0.444
u_{22}	—	—	—	−0.782	−0.776	−0.767	−0.734	−0.736
u_{33}	—	—	—	−0.692	−0.479	−0.491	−1.290	−1.311
u_{44}	—	—	—	−0.593	−0.552	−0.550	−0.694	−0.690
u_{55}	—	—	—	1.815	1.800	1.802	2.237	2.293
u_{12}	—	—	—	—	—	—	0.540	0.378
u_{23}	—	—	—	—	—	—	−0.506	−0.521
u_{34}	—	—	—	—	—	—	−0.388	−0.348
u_{45}	—	—	—	—	—	—	0.345	0.492
u_{21}	—	—	—	—	—	—	0.540	0.648
u_{32}	—	—	—	—	—	—	0.506	0.535
u_{43}	—	—	—	—	—	—	−0.388	−0.449
u_{54}	—	—	—	—	—	—	0.345	0.335

第二个结论是，职业出身和职业获得之间的相关模式大体上具有对称性。将相互对应的对称模型和非对称模型进行比较时发现，只有最后两个模型（即 QSD 和 QD）之间存在显著地偏离对称性特征的情形（$L^2_{QSD}-L^2_{QD}=18.09$；$df=4$；$p<0.01$）。仔细检查这两个模型的 u_{ij} 值发现，最大的不对称性存在于第(1，2)格与第(2，1)格之间。从数据看来，从下层非体力职业阶层向上层非体力职业阶层的向上流动要多于反向的向下流动。u_{ij} 值所体现的另一个可能的不对称性存在于第(4，5)格与第(5，4)格之间，但这或许并不显著，因为第(4，5)格上的样本数很少。正如在第 2 章对称相关部分所提到的，模型中不对称参数的设计只需考虑第(1，2)格与第(2，1)格这一组关系即可。

第2节 | 阶层隔阂与跨越参数

到目前为止,我们尚未论及公式3.1中的参数 c_{ij}。上述所有对角线模型都将 c_{ij} 设定为0。对角线约束模型和对角线非约束模型只是社会距离模型的一种。它们对社会流动过程的刻画主要在于流动所跨越的职业类别的个数。例如,在大多数对角线模型中,每跨越一个界限的流动都由相同的参数($d_{|k|}$ 或 d_k)所决定。因此,同一个参数既适用于从下层体力职业阶层向上层体力职业阶层的流动,也适用于从下层非体力职业阶层向上层非体力职业阶层的流动。

这些过于简单的模型并不能较好地拟合美国的流动表数据。问题在于,流动过程有可能像QD模型所假设的那样复杂。另一方面,问题还可能在于,流动跨越了几个界限与流动跨越了哪些界限相比,其实并不那么重要。假定社会流动是一个人们在行进时不断清除障碍的过程,每清除一个障碍可能类似于跨出同样的一步。对角线模型要说明的是,不管你从哪里开始,第一步的跨度是 $d_0 - d_1$,第二步的跨度是 $d_1 - d_2$,如此类推。相反,也可以假定每一步的跨度是由它所跨越的两个职业类别之间的距离所决定的。在基于这种假定的模型下,上层非体力职业阶层与下层非体力职业阶层之间的障碍跨度是一个值(v_1),不管这是不是第一步,下层

非体力职业阶层与上层体力职业阶层之间的障碍跨度则是另外一个值(v_2),如此等等。根据这种算法,任意两个职业类别 i 和 j 之间的距离就是它们之间的障碍跨度值之和:

$$c_{ij} = \sum_{s=j}^{i-1} v_s \quad (\text{当 } i > j \text{ 时})$$

$$c_{ij} = \sum_{s=j}^{j-1} v_s \quad (\text{当 } i < j \text{ 时})$$

对 c_{ij} 的这一定义与公式 3.1 相同。于是就有了下面的模型,即跨越参数模型(CP):

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + c_{ij} \qquad [3.9]$$

上式适用于公式 2.2。仍以从上层非体力职业阶层向下层非体力职业阶层和从下层非体力职业阶层向上层体力职业阶层的流动为例。根据 CP 模型,上层非体力职业阶层与上层体力职业阶层之间的距离就等于它们之间的障碍跨度值之和。设从上层非体力职业阶层向上层体力职业阶层流动的距离参数为 c_{13},则有 $c_{13} = v_1 + v_2$。CP 模型非常有意思,因为它能形象地展现这种二维空间里的距离感。由于连接任意两个职业类别 i 与 j 的参数等于连接该区间内相邻职业类别的参数之和,所以,所有的职业类别可以说排列在一条直线上。约翰逊(Johnson, 1980:123—130)在美国婚姻配对研究中,将 CP 模型的这一思路进行了进一步的发挥。

CP 模型的另一个有意思的特征在于,职业地位层级两端的继承性被归因于阶层隔阂(即职业类别的这种划分),而不是其他模型所说的本来就存在的阶层继承性。也就是说,CP 模型精确地拟合了第(1, 1)格与第(R, R)格。为了测量

阶层隔阂所没有考虑到的第二类职业一直到第(R－1)类职业的继承性是否显著,又会在CP模型的基础上进一步对参数进行约束,或者相反,不对参数进行约束。在约束形式(QCP-C)下,主对角线上所有的单元格都适用同一个表示继承性强度的参数 d_0:

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + c_{ij} + d_0 \qquad [3.10]$$

上式适用公式2.2和公式3.1。在非约束形式(QCP)下,对角线上的第二个单元格一直到第(R－1)个单元格各自适用一个表示继承性强度的参数 d_{0i}:

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + c_{ij} + d_{0i} \qquad [3.11]$$

上式适用公式2.2,并且仅在 $2 \leqslant i \leqslant R-1$ 时,d_{0i}才存在。哈伯曼(Haberman, 1979:519)曾经给出了一个与此等价的模型,虽然在形式上有所不同。他对任意取值 i 下的 d_{0i}都进行了定义,并通过约束 $v_1 = v_2$ 和 $v_{R-1} = v_R$ 来对 d_{0i}进行区分。与前面所介绍的所有模型相比,CP模型以及它的两个拓展形式在进行参数估计时会更复杂(Goodman, 1972a; Haberman, 1979)。

上述三个模型的统计结果见表3.3。表中的前三栏分别是CP模型、QCP-C模型和QCP模型下的期望频数。第四栏报告了职业类别之间的距离参数(c_{ij})(它由障碍跨度 v_k所决定),对角线以上的部分是在CP模型下得到的估计值,对角线以下的部分则是在QCP模型下得到的估计值[①]。在0.05

① 注意,这段文字中的表述与表格中的内容并不吻合,应该是表格中漏掉了QCP-C模型下的期望频数这一栏。——译者注

表 3.3 父亲的职业与儿子的第一份职业之间的流动：三个跨越参数模型的估计结果

职业出身	职业获得				
	上层非体力	下层非体力	上层体力	下层体力	农民
(1) CP 模型的期望频数					
上层非体力	1414	543.70	287.60	631.54	43.16
下层非体力	664.64	580.66	307.14	674.47	46.09
上层体力	781.96	704.35	776.95	1706.14	116.59
下层体力	847.10	763.02	841.67	3324.06	227.16
农民	413.30	372.27	410.64	1621.79	1832
(2) QCP-C 模型的期望频数					
上层非体力					
下层非体力					
上层体力			缺		
下层体力					
农民					
(3) QCP 模型的期望频数					
上层非体力	1414	566.36	266.26	630.27	43.10
下层非体力	678.64	524	297.64	704.54	48.18
上层体力	744.00	694.09	856	1677.20	114.71
下层体力	849.49	792.80	809.00	3325	227.00
农民	414.87	387.04	395.10	1620.99	1832
(4) 职业类别之间的距离：c_{ij} （对角线以上为 CP 模型结果，对角线以下为 QCP 模型结果）					
上层非体力	—	0.426	0.793	1.086	2.489
下层非体力	0.423	—	0.367	0.660	2.063
上层体力	0.811	0.388	—	0.293	1.696
下层体力	1.086	0.663	0.275	—	1.403
农民	2.488	2.065	1.677	1.402	—

(5) 参数估计值

	跨越参数			对角线参数		
	1	2	3	4	2	3	4

	跨越参数 1	跨越参数 2	跨越参数 3	跨越参数 4	对角线参数 2	对角线参数 3	对角线参数 4
CP	0.426	1.403	0.367	0.293	—	—	—
QCP-C	0.417	1.392	0.365	0.283	0.019	0.019	0.019
QCP	0.423	1.402	0.388	0.275	−0.189	0.189	0.018

	L^2	X^2	df	p	Δ
CP	89.91	90.95	12	<0.01	0.021
QCP-C	89.48	90.65	11	<0.01	0.021
QCP	64.24	64.67	9	<0.01	0.016

的显著性水平上，这三个模型仍然无法拟合数据。不过，如果考虑到Δ值的变化，模型拟合优度的改进还是可以的。CP模型错误地划分了2.1%的个案，QCP模型将这一比例降低到了1.6%。但QCP-C模型则不怎么样——相对于CP模型而言，其拟合优度并没有提高，而且与QCP模型相比明显逊色许多。

需要特别注意的是，CP模型和QCP模型的优势都在于使用了较少的参数。它们在建模时就考虑了上层非体力职业阶层和农民职业阶层所具有的高度继承性。它们之间仅有的实质性区别在于对下层非体力职业阶层和上层体力职业阶层的继承性有不同的考虑。CP模型对下层非体力职业阶层的继承性估计过高，且对上层体力职业阶层的继承性估计过低（请注意，与对角线模型不同，d_{02}的取值在这里为负并不意味着不存在职业地位的继承性）。

两个相邻职业阶层之间的距离的最大值出现在下层体力职业阶层与农民职业阶层之间，其次是在上层非体力职业阶层与下层非体力职业阶层之间。虽然下层非体力职业阶层与上层体力职业阶层之间的流动障碍要稍大于上层体力职业阶层与下层体力职业阶层之间的流动障碍，但两者的绝对值都并不太大。

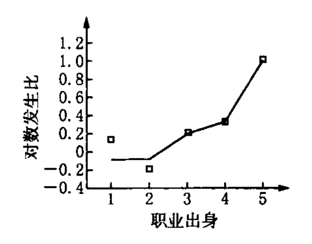
第3节 | 流动模型的 logit 形式

对数线性模型的数学方程式并不一定非要用期望频数来表达。它同样可以用期望 logit 值(定义见下文)或期望优比来表达(Goodman, 1979b)。在 logit 模型形式下,对流动模式的解释通常会更为直观。logit 值就是具有某种相同职业出身(i)的人进入一个较高地位的职业 j 相对于进入一个较低地位的职业(j + 1)的发生比的对数值(简称对数发生比):

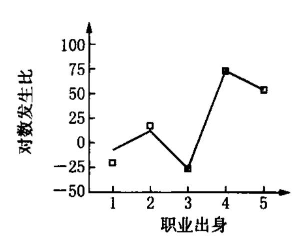
$$\Phi_{ij}^{XY} = \log(F_{ij}/F_{i,\,j+1}) \qquad [3.12]$$

上式中的 X 代表职业出身,Y 代表职业获得。

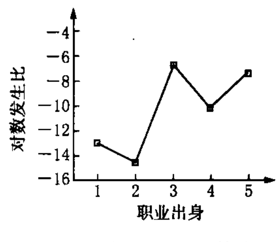
流动模型的 logit 形式比期望频数形式能够更清楚地说明职业出身对职业获得的影响。尽管 logit 模型形式的一般化方程表达经常因太过复杂而很难体现这一点,但每一个单独的 logit 方程却能够很好地说明问题。更重要的是,对这些单独的 logit 模型的结果作图,通常有助于我们理解不同流动模型之间的差异。图 3.2 就是利用 logit 值描绘了美国男性就业者第一份职业获得的模式。图中的点表示实际观察到的 logit 值,线则由 QD 模型下所预期的 logit 值连接而成。该图十分接近并验证了前面所描述的总体流动模式。



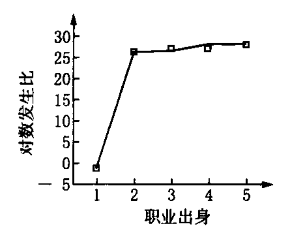
(a) 上层非体力：下层非体力



(b) 下层非体力：上层体力



(c) 上层体力：下层体力



(d) 下层体力：农民

图 3.2　QD 模型下的观测 logit 值(点)与期望 logit 值(线)

其他一些相对简单的模型分别使用不同的方式使图 3.2 中的线变得较为平滑。最简单的模型是完全流动模型，在该模型下，logit 值的连线斜率为 0。QPM 模型只涉及对角线效应，在该模型下，logit 值的连线呈锯齿状。边角模型虽说由对角线效应所产生的峰值所主导，但由于非对角线上的四个边角效应的影响，在其他地方也会出现一些倾斜。QCP 模型不仅有对角线效应所造成的峰值，而且还有一个倾斜模式。这表明，logit 值对地位较高的职业获得比对地位较低的职业获得而言，更多地受到了职业出身的影响。图 3.3 对照了 QCP 模型和 QPM 模型下的 logit 期望值。线是由 QCP 模型下所预期的 logit 值连接而成，点则表示 QPM 模型下预期的 logit 值。

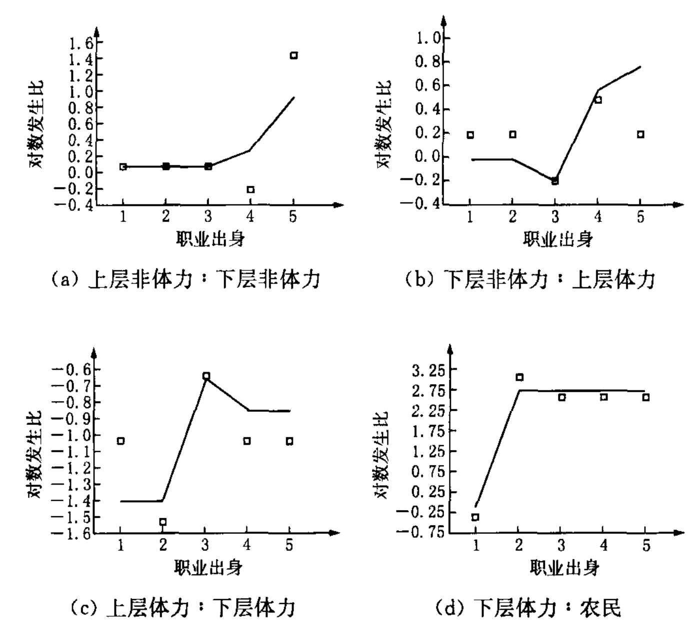


图 3.3 QPM 模型下的期望 logit 值(点)与 QCP 模型下的期望 logit 值(线)

在 QD、QPM 和 QCP 三个模型中，QCP 最好地拟合了数据。因此，到目前为止，该模型下的 logit 值连线图最好地描绘了美国流动表数据中所存在的规律性。其实，模型设计还可以进一步优化，我现在就转向介绍那些与 QCP 模型一样拟合数据，甚至比它拟合得更好的模型。

第4章

拓扑模型

豪斯(1978、1979; Featherman & Hauser, 1978:131—150)重新设计了古德曼(1972a)的一般化模型,以进一步纳入跨行、跨列和跨对角线的参数约束。豪斯从公式3.1中的参数入手发展出了拓扑模型,从而用流动表中的准独立单元格——或"层级"——区域来定义交互效应。在同一层级内,单元格之间是准独立的,也就是说,在同一个层级内,职业出身与职业获得之间是不相关的。同一层级内的每一个单元格上的交互效应都只用一个参数来概括(尽管每一个单元格的参数个数也可以超过一个,Duncan & Schuman, 1980)。同一层级内的所有单元格共享一个交互项参数,于是它们就构成了一个准独立区域。实际上,这个方法就是将模型中的某些参数值约束为相等,如同对角线模型和边角模型的处理手法那样。但与一般化的对角线模型不同,拓扑模型并不事先限定哪些单元格可以被纳入同一个层级。

术语非常重要。本章所涉及的模型通常被称为"结构"模型。这一术语是豪斯(1978)采用的,目的是强调这些模型将普遍效应与交互效应相分离的特性。选择这一术语,同时也是参考了戈德伯格(Goldberger, 1973)的研究,他将那些不管数据形式如何变化,取值总保持不变的参数称为"结构

性的”。豪斯之所以使用这一术语，正是因为这些模型具有不管边缘效应参数怎么变，交互效应都不会变的特点。但是，交互效应参数的这种不变性是有条件的，这会在本章最后一节中加以讨论。拓扑学是几何学的一个分支，涉及的正是变形中的不变性。我将这些模型称为“拓扑”模型，正是要表明它们的不变性是有条件的。

这种统计建模方法是由豪斯(1978、1979)引入的。他用以下文字来概括这一方法(Hauser, 1979:416)：

> 全部单元格(i, j)都被指定分属K个彼此之间完全互斥并已穷尽所有可能的子集，每一个子集共享一个相同的交互项参数(d_k)。因此，除了总体效应、行效应和列效应(即边缘效应)的影响之外，每一个期望频数仅仅由一个交互项参数决定，而这个参数就反映了单元格之间相对流动性或继承性的强度。模型的交互项参数直接与实际观察到的联合强度这一概念相对应(White, 1963:26)，并且可以被解释为职业类别之间的社会距离指标。(比较 Rogoff, 1953:31—32)

子集由一组单元格构成，在同一组单元格里，流动是自由的，即职业出身和职业获得是相互独立的，但这种自由流动或者相互独立是以边缘分布和单元格被指定分属K个层级为条件的。

第1节 | 费瑟曼—豪斯模型

费瑟曼和豪斯(1978:150—159; Hauser, 1979)针对表1.1中的数据提出了有五个层级(即 $K=5$)的拓扑模型。每个单元格被指定的层级序列如以下设计矩阵所示:

$$\begin{matrix} 2 & 4 & 5 & 5 & 5 \\ 3 & 4 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 4 & 4 \\ 5 & 5 & 5 & 4 & 1 \end{matrix}$$

上述矩阵中的数字代表k的取值。注意,层级序列数值从小到大排列,对应的强度(相对于边缘效应而言)则是从高到低变化,两者方向相反。1至5这些数值只不过是在模型参数估计出来以后才赋予的象征性的序列,模型本身并不约束交互项参数以这种整数形式排列:

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_k \qquad [4.1]$$

上式适用公式2.2。该公式只识别层级参数 d_k 之间的差别。也就是说,知道了 $k-1$ 个参数值,就可以推算出第k个参数值。因此,在估计层级参数之前,必须作出某些条件限定。这种模型识别方法带来的后果之一就是模型的自由度个数

为$(R-1)^2-(K-1)$，而并不是一般所认为的$(R-1)^2-K$。公式2.3所体现的标准化方法解决了模型的识别问题，但它并非唯一可行的约束方法。拓扑模型所使用的标准化方法并非这种零和规一，而是约束某一项d_k值为0(即某一个层级的参数并不需要计算)，其他d_k值所测量的正是偏离该基准层级的程度。在实践中，参数值被设定为0的层级就是由最多单元格所组成的子集。这一约束方法有利于比较那些在特定的子集中包含不同数量单元格的模型。而如果用公式2.3那样的约束条件，则每当某一子集k所纳入的单元格数目在模型之间发生变化时，d_k值也必然会随之改变，即使d_k之间的差值(即d_k-d_k)在模型之间并没有发生任何变化。零和规一方法的这种特性正是拓扑模型所采用的标准化方法(即设某一项d_k值为0)所要避免的。本章最后一节会对这一原则加以说明。

K个子集并没有用尽流动表中的准独立区域。在特定的条件下，即使一个2×2分表中的单元格处于不同的层级，它也仍然能够显示出职业出身与职业获得在期望频数分布上的不相关性。以单元格(1，2)，(1，3)，(2，2)，(2，3)组成的这个2×2分表为例，它涉及了第四层级和第五层级，但它根据期望频数得到的优比值在取对数之后为0：

$$\begin{aligned}\theta_{12} &= \log(F_{12}F_{23}/F_{13}F_{23}) \\ &= d_4 + d_5 - d_4 - d_5 \\ &= 0\end{aligned} \qquad [4.2]$$

上述拓扑模型的结果见表4.1。表中所给出的期望频数就是期望频数本身，并非类似之前表格中的$F_{ij}{}^*$。参数的估

计值由 FREQ 软件直接给出，但也可由期望频数推算得到，其方法是要先得到不处于第五层级上的每一个单元格在被假设处于第五层级情况下的期望频数($F5_{ij}^*$)：

$$d_k = \log(F_{ij}/F5_{ij}^*) \qquad [4.3]$$

$F5_{ij}^*$ 的计算原理与 F_{ij}^* 相同。根据定义，只要某个优比涉及处于相同层级上的单元格的期望频数，它的值就总是等于 1。为了得出某个单元格上的 $F5_{ij}^*$ 值，要先找到一个由不处于第五层级上的第(i, j)格和另外三个处于第五层级上的单元格所构成的优比，例如：

$$F5_{22}^* = F_{32}F_{23}/F_{33}$$

或者是找到一个由某些处于第五层级上的单元格和某些 $F5_{ij}^*$ 值已经获知的单元格所构成的优比。

表 4.1 父亲的职业与儿子的第一份职业之间的流动：五层拓扑模型的估计结果

	职业获得				
职业出身	上层非体力	下层非体力	上层体力	下层体力	农民
(1) 期望频数					
上层非体力	1414	515.08	297.43	649.06	44.41
下层非体力	724	522.95	301.97	658.97	45.09
上层体力	754.23	740.04	777.93	1697.62	116.16
下层体力	812.27	796.98	837.79	3328.23	227.73
农民	396.35	388.89	408.80	1624.01	1832
(2) 参数估计值					
上层非体力	1.590	0.599	0	0	0
下层非体力	0.905	0.599	0	0	0
上层体力	0	0	0	0	0
下层体力	0	0	0	0.599	0.599
农民	0	0	0	0.599	3.402

注：$L^2 = 66.57$；$X^2 = 66.50$；$df = 12$；$p < 0.01$；$\Delta = 0.018$。

然而，参数估计结果表明，对角线模型下的参数值所描绘的流动和继承模式正好也被拓扑模型下的这些参数值所反映出来。反映农民职业阶层的继承性的参数值最大，反映上层非体力职业阶层的继承性的参数值次之，随后是从下层非体力职业阶层向上层非体力职业阶层的流动参数。在这三个参数层级中，第三层级是模型矩阵设计中唯一一个不具有对称性的层级。费瑟曼和豪斯(1978:152)指出:“这一不对称性是引人注意的，因为它表示上层白领阶层至少有能力避免某种形式的家庭地位衰落。”这一对称性或许也可以作为某种证据被解读为，代际的向上流动在非体力职业阶层内部较为容易。如果第(1, 2)格的强度——它表示的是上层非体力职业阶层后代的向下流动——异常地低，费瑟曼和豪斯的上述解读会更站得住脚，但数据显示并非如此。事实上，异常的强度值出现在第(2, 1)格，这表明，其实是下层非体力职业阶层后代的向上流动性过高。

流动表边角部分的其余五个单元格上的交互效应在前文的对角线模型和边角模型中被发现十分接近，在这里的拓扑模型中则被约束为完全相等。剩余的 17 个单元格都被标记为具有相对较低的强度(排除了边缘效应的影响之后)，并且在这一大块表格区域(包括整个第三行与第三列上的单元格及其他层级为 5 的单元格)内具有准完全流动的特征，正如第四层级的五个单元格区域内以及第四层级与第五层级之间的流动模式。

费瑟曼和豪斯的这一拓扑模型的拟合优度并不算好。事实上，边角模型、对角线模型以及几个对角线约束模型都具有更低的 L^2 和 X^2 值。但他们更中意这一模型，因为具有

六个层级的候选模型并不能将拟合优度提高多少，而且也提不出具有实质意义的、有趣的单元格子集。尽管如此，为了便于说明，这里还是要探讨一下费瑟曼—豪斯模型的一些改动。在跟随后续的分析过程时，读者需注意，本书所使用的研究样本非常大（N = 19912）。因此，统计模型在能够探测出一些微小效应的同时，也会增加对模型本身的解释难度，毕竟当前的流动理论还有待提高。

第2节 | 模型校正的原则：拟合与简约

流动模型与其他对数线性模型一样，在进行模型选择时，描述性（或拟合优度）和简约性（或自由度）都是很重要的。虽然用完了所有自由度的饱和模型必然会完全复原观测频数的分布，但这种模型描述与简单地考察观测频数的做法相比，并无任何高明之处。正是因为这种简约性与描述性之间的权衡，“一个简单的模型通常比一个较复杂的模型更受青睐，即使后者对数据的拟合更好”（Fienberg，1980：56）。关键在于，以损失拟合优度来换取自由度的做法究竟在多大程度上可以被接受？

模型选择有赖于 L^2 统计量所具有的可叠加性：自由度分别为 df_1 和 df_2 的两个 L^2 统计量之间的差值服从自由度为 df_1-df_2 的卡方分布。在对数线性模型分析中，当一个模型蕴含了另一个模型时，或者说当一个模型对另一个模型的一个或多个参数进行某种约束时，上述原则就适用。约束的形式多种多样，最常见的是零值约束和等值约束。零值约束是将某对数线性模型的一个或多个参数值设定为0；等值约束是将某对数线性模型中的某些参数值设定为相等。如果没有理论约定，上述原则就不能被用于模型选择。过度拟合、

互相参照和不确定结果等问题，都会使纯粹依赖统计技术的模型选择过程变得不可行（Fienberg，1980：56—68）。但即使这些问题真的存在，上述原则仍然提供了有用的指导，而且，它仍然可以被当做模型选择的辅助手段，虽然它不再是可靠且简便的模型选择依据。

费瑟曼—豪斯模型蕴含了边角模型和 QD 模型，因为将后两个模型中的某些参数值约束为 0 或约束为相等就可以得到它。而且，边角模型也包含了 QD 模型。回顾前面的章节就会发现，在屏蔽了边角模型没有加以特别考虑的四个单元格——第(2，3)格、第(3，2)格、第(3，4)格和第(4，3)格——之后，就得到了 QD 模型。这也就是说，边角模型是 QD 模型的一个特例，是将后者中的四个参数（即 u_{23}、u_{32}、u_{34}、u_{43}）都约束为 0。进一步将边角模型中的一个参数约束为 0，5 个参数约束为相等，就得到了费瑟曼—豪斯模型：

$$d_{33} = 0$$

$$d_{12} = d_{22} = d_{44} = d_{45} = d_{54} = d_4$$

请注意，零值约束和等值约束之间的区别有时并不那么确定，因为它是由参数识别的具体标准化方法所决定的。在费瑟曼—豪斯模型的对数形式中，参数的设定可以写成 $d_5 = 0$，$\sum d_k = 0$，如此一来，零值约束其实就变成了等值约束：

$$d_{13} = d_{14} = d_{15} = \cdots = d_{53} = d_5$$

区别两个层级模型的每一个自由度都存在一个约束。因此，两个层级模型之间的区别就可以被分解为$(df_1 - df_2)$个单独检验，每一个检验都有一个自由度。通过加入那些并不显著减小拟合优度的约束条件以及排除那些显著减小拟合优度

的约束条件(所谓的显著是在一些常规的显著性水平上而言,如0.05),模型就可以被校正。非常重要的一点是,刚才提到的显著性水平并不是对找到零假设下 L^2 值这一概率的准确估计,因为在大多数情况下,零假设都是被选来使 L^2 值最大化的。这是典型的过度拟合问题,应该在实践中注意避免。在真实的研究情境下,这里所进行的某些检验可以被用来获得一个可以接受的模型,但是,在进行数量相等的检验之后,用说明性的概率水平来反映真实的概率水平却是一个不好的习惯。

表4.2说明了上述的模型选择过程。QD模型处于起始位置。连续加入四个约束条件之后,QD模型就变成了边角模型。首先,参数 u_{43} 被设定为0,紧接着,参数 u_{34} 也被设定为0,随后依次轮到参数 u_{32} 和参数 u_{23}。这四个约束中的三个都显著地降低了 L^2。在边角模型的对角线参数约束中,唯一在统计意义上不重要的是对参数 u_{34} 的约束。

表4.2所显示的这些约束条件被加入模型中的次序并不是非此不可的。加入次序可达24种之多($4! = 24$)。更重要的是,并不能保证两种不同的加入次序会得出相同的结果。这种情形类似于逐步回归结果的不确定性。因此,约束条件被引入模型的次序会对结果产生影响(Goodman, 1972c)。下文会讲到,对第(3, 3)格的处理方法取决于对第(3, 3)格进行参数约束时是将哪个模型作为基准模型,这正好说明了这一点。

要从边角模型过渡到费瑟曼—豪斯模型,有五个约束条件是必需的。表4.2从唯一的零值约束 $d_{33} = 0$ 入手。将这一约束条件加入边角模型后,L^2 值增加了15.04(见表格下

表 4.2 父亲的职业与儿子的第一份职业之间的流动：连接对角线模型和五层拓扑模型的一系列层级模型

模 型		L^2	X^2	df	p	Δ
M1	对角线模型	15.54	15.48	3	<0.01	0.006
M2	$d_{43}=0$	23.08	22.79	4	<0.01	0.007
M3	$d_{34}=0$	23.25	22.98	5	<0.01	0.007
M4	$d_{32}=0$	43.80	44.12	6	<0.01	0.012
M5	$d_{23}=0$（边角模型）	50.12	50.44	7	<0.01	0.014
M6	$d_{33}=0$	65.16	65.31	8	<0.01	0.017
M7	$d_{12}=d_{22}=d_4$	65.20	65.38	9	<0.01	0.017
M8	$d_{44}=d_4$	65.39	65.45	10	<0.01	0.017
M9	$d_{45}=d_4$	66.22	66.27	11	<0.01	0.017
M10	$d_{54}=d_4$（五层拓扑模型）	66.57	66.50	12	<0.01	0.018
M11	$d_{33}=d_6$	52.10	52.21	11	<0.01	0.015
M12	$d_{32}=d_{23}=d_7$	33.21	33.14	10	<0.01	0.012
M13	$d_{34}=d_{43}=d_8$	21.16	20.90	9	0.01	0.009
M14	$d_{23}=d_{32}=d_{34}=d_{43}=d_7$	29.90	29.69	10	<0.01	0.013
M15	$d_{23}=d_{32}=d_6$；$d_{34}=d_{43}=d_7$；$d_{33}=0$	21.56	21.30	10	0.02	0.009
M16	$d_{23}=d_{32}$；$d_{34}=d_{43}$	20.78	20.50	8	0.01	0.009
模型比较						
C1	M1-M2	7.54	7.31	1	<0.01	
C2	M2-M3	0.17	0.19	1	>0.50	
C3	M3-M4	20.55	21.14	1	<0.01	
C4	M5-M4	6.32	6.32	1	<0.01	
C5	M6-M5	15.04	14.87	1	<0.01	
C6	M7-M6	0.04	0.07	1	>0.50	
C7	M8-M7	0.19	0.07	1	>0.50	
C8	M9-M8	0.83	0.82	1	>0.50	
C9	M10-M9	0.35	0.23	1	>0.50	
C10	M10-M11	14.47	14.29	1	<0.01	
C11	M11-M12	18.89	19.07	1	<0.01	
C12	M12-M13	12.05	12.24	1	<0.01	
C13	M14-M13	8.74	8.79	1	<0.01	
C14	M15-M13	0.40	0.40	1	>0.50	
C15	M15-M16	0.78	0.80	1	>0.50	

半部分以 C5 开头的那一行)，这表明，参数 d_{33} 具有统计显著性(即它的值不等于 0)。其余的四个约束条件针对费瑟曼—豪斯模型中共同构成第四层级的五个单元格。它们中没有一个能显著地提升 L^2 值(见以 C6 至 C9 开头的几行)。因此，从对角线模型向费瑟曼—豪斯模型过渡所加入的诸多约束条件中，有些是相当重要的，有些则不那么重要。重要的部分是约束第(4，3)格、第(3，2)格、第(3，3)格、第(3，4)格为第五层级的一部分。在对角线模型和费瑟曼—豪斯模型之间 51.03 的 L^2 值差异中，这些约束解释了其中的 49.45%。

这一信息可以被用于提高费瑟曼—豪斯模型的拟合优度。然而需要注意，过度拟合的危险此时就变得相当严重了。豪斯(1979:452—426)虽然认为模型拟合优度的提升可能就在于这些单元格，但却出于对过度拟合问题的考虑而没有真正继续下去。在这里，我首先放松的是将第(3，3)格指定为第五层级这一约束，具体方法就是精确地拟合该单元格(也就是将该单元格上的参数层级设定为 6)，这样就得到了模型 M11。它与费瑟曼—豪斯模型相比有显著的改进(见以 C10 开头的那一行)。接下来，我再放松将第(2，3)格和第(3，2)格指定为第五层级这一约束，并加入一个新的层级(第七层级)来指定这两个单元格。正如以 C11 开头的那一行所显示，这一改变显著地减少了 L^2 值。最后，第(3，4)格与第(4，3)格也被从第五层级中分离出来。前面已经讲到，$u_{34}=0$ 这一约束条件并不能显著地改变 L^2 值，但沿袭费瑟曼和豪斯对模型简约性的偏好，第(3，4)格与第(4，3)格被同时引入最后这个约束条件，并且它们是否可以被约束为处于相同

层级这一问题也已进行了检验(见以 C15 开头的那一行)。对第(3, 4)格与第(4, 3)格加入第八个层级之后,就得到了模型 M13,它与模型 M12 相比有显著的改进(见 C12 开头的那一行)。第七层级与第八层级不能被合并为一个相同的层级(合并得到的模型是 M14),正如以 C13 开头的那一行所显示。然而,既然围绕次对角线的单元格现在都已被排除在第五层级之外,那么第(3, 3)格就可以被恢复为第五层级了(见模型 M15),正如以 C14 开头的那一行所显示的。注意,模型 M15 与 M13 而非 M14 之间存在层级关系。最后一步是检验第七层级和第八层级的对称性。两个检验是分别进行的,但表中只给出了合并后的结果。若将这两个约束条件加在一起,L^2 值会减少 0.78(见以 C15 开头的那一行)。

模型 M15 的结果见表 4.3。将表中的参数估计值与 QD 模型下的参数估计值(见表 3.2)比较一下可以发现,两个模型所揭示的流动模式几乎相同。模型 M15 优于 QD 模型之处在于它更简约,它清理了 QD 模型中差别不大的几个参数(模型 M15 中有六个等值约束),并删去了一个并不重要的参数(d_{33})。与费瑟曼—豪斯模型的假设相比,模型 M15 揭示出了更少的跨越体力/非体力界限的流动($d_7 = -0.248$)。这一点其实在 QD 模型中也已经显示出来了,但对角线模型并没有清楚地表明,这种跨越体力/非体力界限的流动的缺乏在两个相反的方向上(即向上流动和向下流动)是等量的。流动性的缺乏还出现在上层体力职业阶层与下层体力职业阶层之间,但在量上有所改善。当然,模型 M15 的 p 值(0.02)小于常规的显著性水平。如果加入新的层级参数,模型的拟合优度或许还可以再提高。但这种做法并不可取。

因为，此时数据中的随机误差被放大的可能性已经很大了。

表 4.3 父亲的职业与儿子的第一份职业之间的流动：七层拓扑模型(M15)的估计结果

职业出身	职业获得 上层非体力	下层非体力	上层体力	下层体力	农民
(1) 期望频数					
上层非体力	1414	508.60	301.38	652.96	43.06
下层非体力	724	540.12	249.73	693.42	45.73
上层体力	781.79	652.27	865.35	1662.95	123.64
下层体力	798.60	853.92	784.05	3345.78	220.65
农民	382.59	409.09	423.48	1602.86	1832
(2) 参数估计值					
上层非体力	1.647	0.558	0	0	0
下层非体力	0.918	0.558	−0.248	0	0
上层体力	0	−0.248	0	−0.120	0
下层体力	0	0	−0.120	0.558	0.558
农民	0	0	0	0.558	3.410

注：$L^2 = 21.56$；$X^2 = 21.30$；$df = 10$；$p < 0.02$；$\Delta = 0.009$。

如果我没有发现跨越体力/非体力界限的流动性较弱这一有趣的现象，我的确会倾向于费瑟曼和豪斯关于五个层级的拓扑模型就足够了的说法。但我现在当然不会同意，尤其是当我注意到，费瑟曼和豪斯(1978:166—172)关于父代职业与子代现职的流动分析结果也同样显示出跨越体力/非体力界限的向下流动(但并不包括相应的向上流动)有所缺乏。七层拓扑模型表明，社会中层出身的男性的流动机会比费瑟曼和豪斯(1978:178)的五层拓扑模型所预期的要少。

五层拓扑模型和由此延伸得到的七层拓扑模型的 logit 形式结果见图 4.1。图中的点代表五层模型的期望 logit 值，线则连接了七层模型的期望 logit 值。这些图清楚地展现了

七层模型如何优于五层模型。请留意图 4.1 中五层拓扑模型的 logit 值与图 3.3 中准跨越参数模型的 logit 值之间的相似性。这一现象是由彭蒂宁（Pontinen，1981）发现并指出的。至于这种相似性因何产生，又会造成何种结果是下一节我们要讨论的主题。

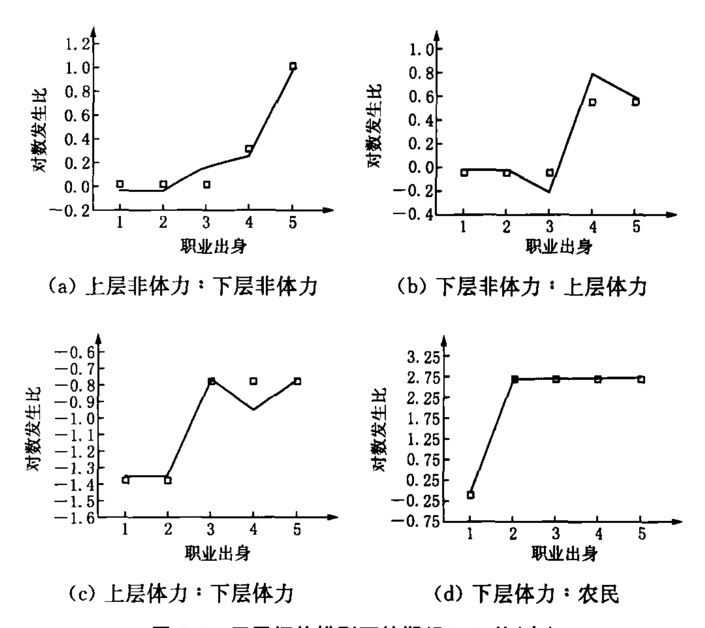


图 4.1 五层拓扑模型下的期望 logit 值(点)与七层拓扑模型下的期望 logit 值(线)

第3节 | 拓扑模型的不确定性

大多数对数线性模型都有一种以上的写法。利用古德曼(1972a)提供的公式,能够得到本书到目前为止所介绍的大部分模型的等价方程式。某一模型的等价方程式通常在意义解释上也是等价的,但却并非总是如此。试想一下彭蒂宁(1981)提出的基于费瑟曼—豪斯五层模型(本节中简称为FH5)的两种变形:

2	4	4	5	5
3	4	4	5	5
5	5	4	5	5
5	5	4	4	4
5	5	4	4	1

P1

2	4	5	5	5
3	4	5	5	5
4	4	4	4	4
5	5	5	4	4
5	5	5	4	1

P2

将FH5中整个第三列上的单元格从第五层级改为第四层级,就得到P1。同样,将FH5中整个第三行上的单元格从第五层级改为第四层级,就得到P2。FH5、P1和P2看似不同,但其实相同。三个模型的期望频数分布完全相同。

FH5、P1和P2之间的差异意味着对它们的解释也不同。FH5的一个重要特征是,第五层级所涉及的流动表区域

不仅广阔而且具有对称性。费瑟曼和豪斯(1978:151—153)在解释 FH5 时,对这两个特征都做了强调。FH5 所具有的广阔对称性并非某种巧合。费瑟曼和豪斯其实是特意挑选了这一具有最大程度对称性的模型(Featherman & Hauser, 1978:150—176; Hauser, 1978、1979、1981)。这种对称性对于他们拒绝布劳和邓肯(1967:58—67)的相关结论来说显得尤其重要,后者曾总结道,具有不完全可渗透性的阶层界限在允许从体力职业阶层到非体力职业阶层以及从农民职业阶层到体力职业阶层这种向上流动的同时,也阻碍了跨越这些阶层界限的向下流动。P1 和 P2 包含了 FH5 的部分对称结构,但却并不包含这一关键的对称特征,即发生于上层体力职业阶层与下层非体力职业阶层之间的双向流动(即向上流动与向下流动)具有对称性。在 P1 中,从下层非体力职业阶层到上层体力职业阶层的向下流动多于反向的向上流动。在 P2 中,跨越体力/非体力界限的向上流动多于相应的向下流动,而这恰好与布劳和邓肯的上述结论相一致。因此,不能将 FH5 对表 1.1 流动表数据的拟合用来检验布劳和邓肯关于阶层界限具有不完全可渗透性这一结论,原因在于,与布劳和邓肯这一结论相等价的模型其实本来就有悖于 FH5 的对称特性。

FH5、P1 和 P2 之间的关系,可以由表 4.4 所报告的参数估计值反映出来。最引人注意的是,三个模型的交互项参数彼此完全相同。此外,边缘效应也十分接近。FH5 与 P1 的行边缘参数完全相同,FH5 与 P2 的列边缘参数完全相同。FH5 与 P1 之间只有第三列的边缘参数不同,FH5 与 P2 之间只有第三行的边缘参数不同。P2 第三行的参数值比 FH5

和 P1 的相应参数值小 0.599，P1 第三列的参数值比 FH5 和 P1 的相应参数值也小 0.599，而第四层级上的交互效应参数也是 0.599，这并非巧合。虽然彭蒂宁（1981：6）经过一系列相当复杂的运算之后才得到了 P1（和 P2），但这一运算过程其实就相当于向第三列（对 P2 而言是第三行）加入第四层级上的参数，然后再与那一列（对 P2 而言是行）的边缘效应相减。期望频数的对数值（以及期望 logit 值和优比值）不会发生任何改变，因为加到交互效应上的值又被从列（或行）边缘效应中减去了。

P1 和 P2 还不是仅有的两个能够得到与 FH5 相同的期望频数的模型变形。麦克唐纳（Macdonald，1981）就提出了这样一个拓扑模型：

2	2	2	3	3
4	2	2	3	3
5	3	2	3	3
5	3	2	2	2
5	3	2	2	1

M1

M1 是从 P1 得到的，方法是将第一列的层级参数减去 0.99（即 d_1-d_4），并将 a_{21} 加上 0.99。表 4.4 也给出了 M1 的参数估计值（注意，这些参数值是相对于第三层级而言的），这样就可以将其与 FH5、P1 和 P2 进行比较。这种形式（而非 Macdonald 所用的形式）的比较突出了从 P1 得到 M1 或从 FH5 得到 M1 的运算方法。

这些结果确实带来了麻烦，但还不至于麻烦到像麦克唐纳（1981）所说的拓扑模型应该被舍弃的程度。这些结果强

表 4.4 父亲的职业与儿子的第一份职业之间的流动：各种五层拓扑模型的参数估计值比较

效 应		FH5	P1	P2	M1	CP*
行效应	1	−0.318	−0.318	−0.318	−0.318	0.315
	2	−0.303	−0.303	−0.303	−0.303	0.119
	3	0.643	0.643	0.044	0.643	−0.432
	4	0.718	0.718	0.718	0.718	−0.904
	5	—	—	—	—	—
列效应	1	1.872	1.872	1.872	2.862	0.965
	2	1.853	1.853	1.853	1.853	0.451
	3	1.903	1.304	1.903	1.304	0.201
	4	2.683	2.683	2.683	2.683	1.281
	5	—	—	—	—	
交互效应	1	3.403	3.403	3.403	3.403	0.495
	2	1.590	1.590	1.590	0.599	0.300
	3	0.905	0.905	0.905	—	0.300
	4	0.599	0.599	0.599	−0.086	1.402
	5	—	—	—	−0.989	0.306

调了研究者必须小心、恰当地解释拓扑模型的参数结果。值得注意的是，FH5、P1、P2 和 M1 这四个等价模型中的三个，在相同列单元格上的层级参数差异并不受模型变形的影响。同样，其中两个模型在相同行单元格上的层级参数差异也不受模型变形的影响。

尽管上述四个等价模型都意味着相同的观测频数分布，但它们的实质含义还是可以并且确实不同的。这并不意味着对数线性模型不适合用来分析流动表数据，也不意味着研究者有理由无视那些能够表明某个模型优于另一个模型的证据。豪斯（1981）注意到，研究者在用结构方程模型来分析协方差矩阵时所面临的选择性问题与这种情形十分相似。例如，某研究者可能对高中学生的成就动机与学习成绩之间的关系感兴趣。一旦这种相关关系成立，就可以提出四个模

型:可能是动机决定成就,可能是成就决定动机;动机和成就可能相互决定,动机和成就之间也可能是由测量误差所引起的伪相关。四个模型意味着同样的协方差矩阵,但我们仍然承认对它们进行继续分析的重要性,并必须决定究竟选择哪种模型。所以,这与对数线性模型下的情形是相同的。

拟合优度不能被用来区别上述四个模型。由于这些模型能够得到相同的期望频数分布,所以它们的拟合优度也是相同的。但是,拟合优度并非模型选择的唯一准则。豪斯(1981)将流动模型是否具有对称性作为另一个可取的准则,他表明,FH5 所包含的对称性结构不仅是必需的,而且能充分体现美国流动表数据偏离准对称模型(见第 2 章有关对称相关的部分)这一理想模型的程度。

流动模型或许还可以在重复研究过程中加以区别。当进行历时比较、跨国比较或不同亚群体之间的比较时,等价模型会体现出差异性。例如,假定 P1"真实地"描述了从父代职业向子代第一份职业的流动情况。如果在 P1 中,第四层级与第五层级之间的流动障碍缩减为 0(即第四层级中所有单元格上的交互效应都减少 0.599),将只有 P1 能以最简单的参数改动反映这种变化,即只需改动一个交互项参数。所有其他模型都会涉及外在边缘参数的变化以及交互项参数的一些变化(后一种情况是对 M1 而言)。按照简约性的要求,应该选定这种参数最少发生变动的模型。因此,在上述假定情形下,我们会选择 P1 作为最优模型。当然,真实情形下的重复研究结果并不一定能像刚才的假定情形那样清楚。但是,历时比较、跨国比较和不同亚群体之间的比较仍然可以为模型选择提供重要的线索。第 5 章关于等价模型的部

分会再举例说明如何使用跨国比较的研究结果来进行模型选择。

在进入其他主题之前，关于模型的等价性还要最后补充一点，即拓扑模型与其他模型之间也可能存在等价性。彭蒂宁就发展出了这样一个非常有意思的模型，姑且称之为"跨越参数改动模型"，它与 FH5 是等价的。该模型将第二个和第三个跨越参数约束为相等（$v_2 = v_3$），并且特别为第(2, 1)格引进了一个参数。该模型（简写为 CP^*）的参数估计结果见表 4.4 中的最后一列。表中所给出的第一到第四个交互项参数就是前四个 v_i 的估计值，第五个交互项参数则是第(2, 1)格的参数估计值。这一结果与其他等价模型的结果大相径庭。首先，拓扑模型与跨越参数模型所强调的实际流动过程不同。拓扑模型更具有不连续性，它关注的是没有排序、不被约束的交互项参数的聚类。跨越模型关注的则是阶层隔阂，它取决于职业的分类排序和公式 3.11 中的约束条件。两类模型都强调了短距离流动多于长距离流动的事实，但即使在这一点上，它们仍有所不同。即使流动表数据中存在过多的长距离流动，拓扑模型应该也能够轻易地拟合数据，但跨越模型却只能拟合存在过多短距离流动的数据。

第5章

尺度相关模型

流动表分析与社会经济地位获得研究一样，探讨的理论议题都在于代际效应(例如，Blau & Duncan, 1967; Featherman & Hauser, 1978)。但是，豪斯(1978)和其他研究者(Pullum, 1975:2; Duncan, 1979)都注意到，这两个研究传统泾渭分明，明显地表现在它们利用数据方式的不同。从本书前面的部分不难看出，流动表分析的重点在于找到流动表所含不同分表中存在的或强或弱的关联模式。与此不同，社会经济地位获得研究着重的是社会出身与地位获得之间的因果关系的强度。这种差别可以从两个学派的方法论上反映出来。应用于表格数据的对数线性模型满足了流动表分析对所有单元格进行逐一审视的要求，而一般线性模型更好地适用于地位获得研究所采用的多变量分析路径。

在度量职业特性这一议题上，地位获得研究与流动表分析之间也有所区别。大多数的流动表分析都很少对测量尺度作出某些假定(尽管有一些模型确实也会对职业类别之间的排列次序作出假定)。对地位获得研究而言，选择一个合适的测量尺度则非常必要，因为它关注的正是在同样的职业声望、地位或权威尺度上，父代位置与子代位置之间所具有的关系。虽然邓肯(1979)对这两种研究路径作出了某种调

和，即先将西蒙(Simon，1974)的模型应用于流动表，随后再发展出一个增加了约束条件的模型，但对这些模型进行扩展仍然是古德曼(1979a)的贡献。

第1节 | 统一关联

研究社会流动的地位获得模型假定社会出身与地位获得之间存在一种线性关系。这个模型认为，社会出身每增加一个单位值，地位获得就会相应地增加某个单位值（再加上或减去一个服从正态分布的误差项）。假定这一模型是正确的，再进一步假定地位尺度被均分为 R 个间隔，根据这一尺度测量的社会出身和地位获得在进行交叉分类后得到的 $R \times R$ 表中，第一行/第一列（$i = 1$；$j = 1$）代表最高地位，第 R 行/第 R 列（$i=R$；$j=R$）代表最低地位。在上述条件下，给定出身类别 i 时，个体获得某一地位类别(j)相对于获得稍低一级地位类别（$j+1$）的发生比为：

$$\begin{aligned}\Phi_{ij} &= \log(F_{ij}/F_{i,\,j+1}) \\ &= a_{2j} - a_{2,\,j+1} + b_i \qquad [5.1]\end{aligned}$$

上式适用于公式 2.2，b 是一个统一关联参数（在特定条件下等于回归系数）。统一关联与回归之间的相似性如图 5.1 所示，该图给出了美国流动表数据的实际 logit 值以及用统一关联模型进行拟合之后所得到的期望 logit 值（实际 logit 值用点表示，期望 logit 值用线表示）。

请注意，用统一关联模型得到的 logit 值与社会出身变量

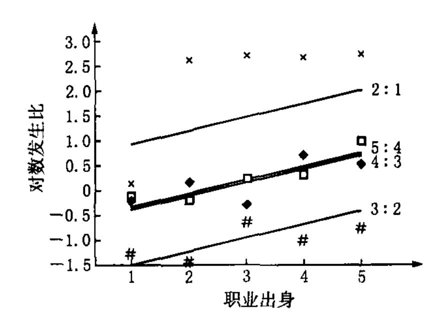


图 5.1 统一关联模型下的观测 logit 值(点)与期望 logit 值(线)

之间具有线性关系，即每当社会出身增加一个单位值，它也增加一个固定的单位值。这些直线是平行的，斜率(b)都为 0.269。统一关联与回归之间还有其他方面的相似性(Haberman，1979:396；Logan，1983)，但这超出了本章的范围，恕不赘述。统一关联模型对美国流动表数据的拟合并不好。有关如何提高其拟合优度的方法，会在介绍完该模型的几种数学表达方式之后再讨论，但图 5.1 已经明显反映出继承性是最大的问题所在。

在统一关联模型下，期望频数服从下述公式：

$$\log(F_{ij}) = a_0 + a_{1j} + a_{2j} + b_{ij} \qquad [5.2]$$

上式适用于公式 2.2。统一关联模型这一命名来自它所预期的对数优比值具有如下模式：

$$\begin{aligned}\theta_{ij} &= \log(F_{ij}F_{i+1,\,j+1}/F_{i+1,\,j}F_{i,\,j+1}) \\ &= b\end{aligned} \qquad [5.3]$$

也就是说，社会出身与地位获得之间的相关性在构成流动表的每一个 2×2 基本单元分表中都是相同的。统一关联模型是哈伯曼(1974)的线性乘线性交互模型的一种特殊形式：

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + bX_iX_j \qquad [5.4]$$

X_i 和 X_j 分别表示行变量和列变量各个类别的度量得分。令线性乘线性交互模型中的 X_i 和 X_j 相同并等于一组间距相等的常数，就得到了统一关联模型。这一约束可以表达为：

$$X_i = (R+1) - i,\ X_j = (R+1) - j \qquad [5.5]$$

根据社会经济地位的相对高低，职业类别在这里就被度量为从 5 到 1。

第2节 | 统一关联的一般化

图5.1可以被用于诊断统一关联模型在拟合美国流动表数据时所存在的问题。图中有四处可以校正，改动其中的任何一处或许就可以提高模型的拟合优度。第一，将logit值约束为一条直线可能过于苛刻；第二，每条线的斜率可能并不完全相同；第三，实际logit值与期望logit值出现最大偏差的地方在于代际继承发生时，这表明需要将对角线单元格进行特殊处理；最后，农民职业出身的人的流动模式最为不同。古德曼(1979a)提出了一个模型，分别针对上述三处进行模型校正。对第四处进行校正的模型方案更为直截了当，例如，将农民职业出身的人从样本中删除(Blau & Duncan, 1967: 135)。

行效应模型(Simon, 1974; Duncan, 1979; Goodman, 1979a)将获得一种地位类别相对于获得另一种地位类别的logit值与社会出身之间的关系表述为一个非线性函数，但约束这种关系对所有logit值来说都相同。换言之，行效应模型是指，社会出身每改变一个类别，获得某一地位类别相对于获得稍低一级地位类别的发生比的对数值都改变(增加或减少)一个固定值，不管比较的是哪两个相邻的地位获得类别。行效应模型的logit形式为：

$$\Phi_{ij} = \log(F_{ij}/F_{i,\,j+1})$$
$$= a_0 + a_{2j} - a_{2,\,j+1} + b_0 X_i + b_{1i} \qquad [5.6]$$

上式适用于公式 2.2 和公式 5.5，并且 $b_{11} = b_{1R} = 0$。对 b_{11} 和 b_{1R} 的这一约束是为了解决模型识别的问题(Goodman, 1979a)。其他形式的约束也是可行的(Clogg, 1982b)，但我偏向于这一约束，因为 b_{12} 一直到 $b_{1,\,R-1}$ 的参数估计结果在解释时具有几何学上的意义。将 Φ_{1j} 到 $\Phi_{R-1,\,j}$ 连成一条线，并将其标注为 AB_j，则 AB_j 的斜率将会等于 b_0。b_{12} 到 $b_{1,\,R-1}$ 测量的是 Φ_{2j} 到 $\Phi_{R-1,\,j}$ 这些参数值偏离 AB_j 的程度。注意，虽然行效应模型允许期望的 logit 值偏离直线 AB_j，但并不约束 AB_j 之间是平行的，也不约束对 AB_j 的偏离程度(即 b_{1i})对所有的 j 来说是相同的。因此，连线上所有相邻两点之间是等距的。结果见图 5.2(a)。

列效应模型(Goodman, 1979a)放松了对斜率的对束，但保留了线性特征，正如图 5.2(b)所示。这一模型包含了能够改变期望 logit 值连线的斜率的参数：

$$\Phi_{ij} = a_{2j} - a_{2,\,j+1} + (b_0 + b_{2j} - b_{2,\,j+1})X_i \qquad [5.7]$$

上式适用于公式 2.2 和公式 5.5，并且 $b_{21} = b_{2R} = 0$。像行效应模型一样，出于模型识别的原因，参数中的两个极值被设定为 0。但在这里，采用这一约束的原因是为了对称性，而不是因为这些系数解释起来更为直观。

行效应和列效应可以通过相加的方法结合起来，于是又得到了一个模型(Goodman 称之为“模型 Ⅰ”)。这一模型既允许对线性特征的偏离，也允许连接 1_j 到 R_j 的连线 AB_j 的斜率可变，线性偏离程度通过 b_{1i}(i 的取值从 2 到 R－1)来测

量，斜率由 $b_{2j}-b_{2,j+1}$（j的取值同样是从2到R－1）这一差值来测量。行与列效应模型（模型Ⅰ）的数学表达式为：

$$\Phi_{ij}=a_{2j}-a_{2,j+1}+(b_0+b_{2j}-b_{2,j+1})X_i+b_{1i} \quad [5.8]$$

上式适用于公式2.2和公式5.5，并且 $b_{11}=b_{1R}=b_{21}=b_{2R}=0$。图5.2(c)展示了用该模型来拟合美国流动表数据后所得到的期望logit值结果。

有一种可能性很有意思，就是行效应与列效应相等，即 $b_{1k}=b_{2k}\ (i=k,\ j=k,\ 2\leqslant k\leqslant R-1)$。这一模型在其他表格数据（例如，由丈夫职业和妻子职业构成的列联表）建模中的重要性比其在代际流动分析中的重要性更高（Hout, 1982）。它就是大家知道的同质行与列效应模型（模型Ⅰ）.

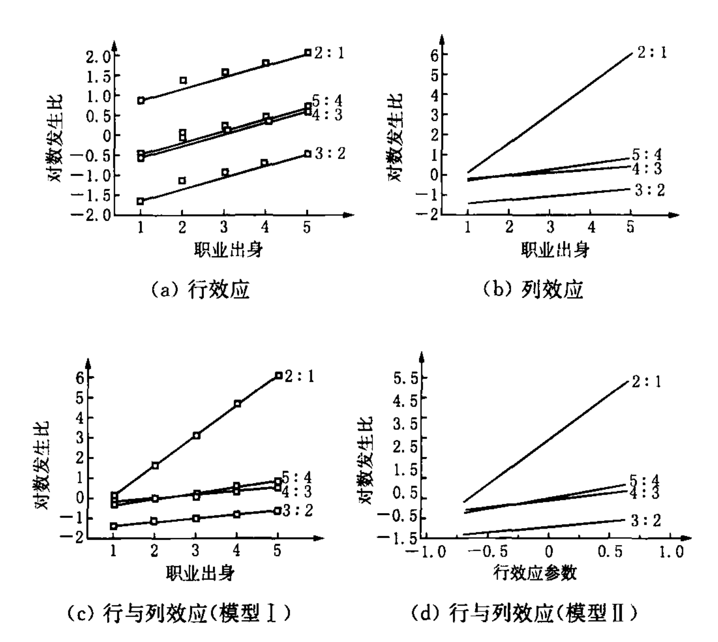


图5.2 各种关联模型下的期望logit值

第 3 节 | 模型Ⅱ

古德曼(1979a)和克洛格(1982b、1983)从另一个不同的角度探讨了如何校正统一关联模型的参数约束。与关注引入非线性特征或斜率变化这种思路有所不同,他们设法分别或者同时对社会出身和地位获得进行尺度上的重新度量,以得出线性乘线性交互参数,而这些参数则会形成一组平行直线。古德曼(1979a)将公式 5.8 所表达的模型命名为"模型Ⅰ",并提出了它的一个替代模型(命名为"模型Ⅱ"):

$$\Phi_{ij} = a_{2j} - a_{2,j+1} + bu_i(v_j - v_{j+1}) \qquad [5.9]$$

并且,

$$\sum u_i = \sum v_j = 0$$
$$\sum u_i^2 = \sum u_j^2 = 1 \qquad [5.10]$$

在这一模型中,u_i 是一组针对社会出身变量的尺度调整参数,与之相应,v_j 是一组针对地位获得变量的尺度调整参数。这两组参数值都需要从数据中估计出来。该模型是线性乘线性交互模型的一种形式,即将不同职业类别的地位得分作为需要估计的参数,而不是由理论或前人的研究结果来决定。如果各种职业类别的地位得分是未知的,那么,模型Ⅱ

会对这些得分进行估计，以得到能最佳拟合数据的线性乘线性交互效应。如果 u_i 和 v_j 被发现是均等间隔的（即对所有的 i 值来说，$u_i - u_{i+1}$ 总是等于某个常数，对所有的 j 值来说，$v_j - v_{j+1}$ 也都等于这个常数），模型Ⅱ就等同于统一关联模型。如果只有 u_i 是均等间隔的，那么模型Ⅱ就等同于列效应模型（公式 5.7）；如果只有 v_j 是均等间隔的，那么，模型Ⅱ就等同于行效应模型（公式 5.6）。用模型Ⅱ拟合美国流动表数据的结果见图 5.2(d)。该图最明显的特征是 X 轴的尺度发生了变化，它反映的就是 u_i。图中的线并不是平行的。参数 v_j 带来的尺度变化可以被应用于 ij，以使这些线相互平行。注意，模型Ⅱ并不具有对数线性的特征，因为 b、u_i、v_j 都必须被估计出来。至于估计方法具体如何，有兴趣的读者可以参考其他著作（Goodman，1979a；Clogg，1983）。

模型Ⅱ可以将行效应和列效应约束为相等，与此类似，模型Ⅱ也可以有这种同质效应形式。这种约束形式下的模型Ⅱ具有特别重要的实际意义，因为它假定社会出身和地位获得的测量尺度具有一致性，即 $u_k = v_k (i = k,\ j = k,\ 2 \leqslant k \leqslant R-1)$。

到目前为止，我们尚未谈及自由度的问题。事实上，这些模型相当简约，它们没有耗费很多自由度就概括了社会出身与地位获得之间的相关模式。这种简约性在大表中表现得尤为明显（例如，Breiger，1981；Hout，待出）。统一关联模型是最简约的，它在拟合边缘分布之外只需再拟合一个参数。行效应模型和列效应模型都需要拟合除边缘分布以外的 R－2 个参数，两个同质行与列效应模型也是如此。模型Ⅰ和模型Ⅱ需要拟合的参数最多，有 2(R－2)个，但尽管如

此，它还是相当简约的。将拟合了边缘分布之后剩余的自由度个数$(R-1)^2$再减去独立参数的个数，就得到了各个模型的自由度：

模　　型	自由度
统一关联模型	$(R-1)^2-1$
行效应模型	$(R-1)(R-2)$
列效应模型	$(R-1)(R-2)$
行与列效应模型(Ⅰ)	$(R-2)^2$
同质行与列效应模型(Ⅰ)	$(R-1)(R-2)$
行与列效应模型(Ⅱ)	$(R-2)^2$
同质行与列效应模型(Ⅱ)	$(R-1)(R-2)$

第4节 | 对美国流动表数据中相关性的分析

表5.1简要地给出了对美国流动表数据中相关性的分析结果。完全流动模型——零相关模型——也列在表中以进行比较。表中没有任何一个模型能够充分拟合数据，但这并不代表这些模型不能提供有用的信息。相反，它们用很少的参数说明了表格中大部分的相关信息。当引入单一的统一关联参数时，模型的拟合优度得到了最大程度的提升。统一关联模型的 L^2 值比零相关模型的 L^2 值减少了2280.69。行效应与列效应都能显著地提高模型的拟合优度，但行效应参数与列效应参数并不相等，正如同质行—列效应模型与异质行—列效应模型的 L^2 值之差所显示的那样。

这些模型的参数估计值见表5.2。由于模型对数据的拟合并不好，所以这些参数值并不具备太大的实质性意义，但这里还是可以谈谈对它们的理论解释。统一关联模型和行效应模型中的统一关联参数表示的是连接 Φ_{ij} 和 Φ_{Rj} 的直线的斜率。在列效应模型和模型Ⅰ中，统一关联参数表示的是斜率的基准值，列效应会使斜率发生变化；在模型Ⅱ中，统一关联参数表示的是在进行尺度调整之后连接 Φ_{ij} 和 Φ_{Rj} 的直线的斜率。为了识别模型Ⅱ，需要对它进行一些约束。

表 5.1 父亲的职业与儿子的第一份职业之间的流动：关联模型的估计结果

	职业获得				
职业出身	上层非体力	下层非体力	上层体力	下层体力	农民
(1) UA 模型的期望频数					
上层非体力	1407.07	647.02	320.36	483.31	62.24
下层非体力	778.96	468.87	303.89	600.12	101.16
上层体力	894.95	705.14	598.23	1546.45	341.23
下层体力	734.60	757.64	841.38	2847.06	822.33
农民	285.42	385.33	560.14	2481.07	938.05
(2) 模型 I 的期望频数					
上层非体力	1348.01	579.03	345.33	646.21	1.42
下层非体力	746.52	437.43	319.05	742.53	7.47
上层体力	842.18	686.74	624.92	1845.22	86.94
下层体力	864.02	864.14	864.69	2855.13	555.02
农民	300.26	396.67	470.01	1868.91	1614.14
(3) 模型 II 的期望频数					
上层非体力	1311.81	564.52	331.18	708.98	3.47
下层非体力	700.94	426.31	313.36	795.50	16.88
上层体力	862.63	695.04	613.66	1789.46	125.21
下层体力	987.70	923.79	898.88	2821.03	371.60
农民	237.91	354.34	466.93	1843.04	1747.79
	L^2	X^2	df	p	Δ
零相关模型	6170.13	7166.77	16	<0.05	0.205
统一关联模型	2280.69	2220.56	15	<0.05	0.120
行效应模型	2080.17	2042.67	12	<0.05	0.123
列效应模型	903.84	1680.36	12	<0.05	0.069
模型 I	877.81	1824.29	9	<0.05	0.067
模型 II	685.51	990.02	9	<0.05	0.065
同质模型 I	1287.89	1343.50	12	<0.05	0.089
同质模型 II	935.87	1056.91	12	<0.05	0.083

表 5.2 父亲的职业与儿子的第一份职业之间的流动：各种关联模型的参数估计值(括号内的数字为标准误)

参数	统一关联模型	行效应模型	列效应模型	行与列效应模型			同质行与列效应模型	
				I	II(a)	II(b)	I	II(b)
统一关联效应	0.269*	0.295*	0.525*	0.534*	4.574	0.269	0.389*	0.269
	(0.005)	(0.006)	(0.011)	(0.011)			(0.007)	
行效应								
1	—	0	—	0	0.642	5	0	5
2	—	0.064*	—	−0.030	0.271	3.469	0.032*	3.950
		(0.020)		(0.021)			(0.005)	
3	—	0.113*	—	−0.079*	−0.031	2.225	0.155*	3.178
		(0.016)		0.018			(0.009)	
4	—	0.221*	—	−0.003	−0.191	1.565	0.321*	2.736
		(0.016)		0.017			(0.011)	
5	—	0	—	0	−0.691	−0.493	0	0.065
列效应								
1	—	—	0	0	0.456	5	0	5
2	—	—	0.236*	0.253*	0.252	4.161	0.032*	3.950
			(0.019)	(0.021)			(0.005)	
3	—	—	0.593*	0.615*	0.119	3.613	0.155*	3.178
			(0.026)	(0.029)			(0.009)	
4	—	—	0.940*	0.960*	0.019	3.200	0.321*	2.736
			(0.031)	(0.035)			(0.011)	
5	—	—	0	0	−0.845	−0.362	0	0.065

注：对模型II无法计算标准误。* 表示 $p < 0.05$。

表 5.2 给出了两种标准化方法。模型Ⅱ(a)采用的是前面讲过的方法(公式 5.10)。模型Ⅱ(b)采用的方法是约束该模型下的统一关联参数等于统一关联模型下的统一关联参数，并且约束上层非体力职业阶层的行效应与列效应都等于 5.0。后一种标准化方法强调将模型Ⅱ理解为是对统一关联模型的测量尺度加以调整。行效应与列效应可以被解释为利用经验数据得出的行与列的尺度得分。如果这些行系数与列系数是事先已知的，它们就可以代替(5，4，3，2，1)作为行与列的尺度得分。使用这些行与列的尺度得分来拟合线性乘线性交互模型，会得到与模型Ⅱ同样的 L^2 值以及一个统一关联参数 0.269。既然行与列上的职业分类相同，理想的做法就是对行与列使用同样的尺度得分，但如果不能显著地减少 L^2，就表明这种做法不妥，因为具有这种同质性约束的模型Ⅱ并不能达到没有这种同质性约束的模型Ⅱ的拟合优度。

同质性的不成立，可能是由那些导致模型Ⅱ拟合度不好的外在因素的干扰引起的。但如果模型能够拟合数据，并且行与列的尺度得分仍然不同，我们可以下结论说，不同职业之间的距离已经在代际之间发生了变化，或者说，职业分类体系的构成在父代与子代之间已经不再相同。就美国流动表数据而言，我会倾向于接受第二种结论，因为美国劳动力结构的重大变化确实已经导致了父代与子代在职业分布上的一系列差异，相信这种变化已经在使用宽泛的职业五分法的 5×5 流动表中体现出来了。

第5节 删除对角线单元格后的相关性分析

提高上述模型的拟合优度的方法仍然是像之前做过的那样:从对角线入手。正如其他对数线性模型那样,通过屏蔽某些特定的单元格或者对这些单元格拟合某些独特的参数(两者是等价的),就可以对本章所介绍的各种关联模型进行修正。以后一种方法为例,准统一关联模型的表达式为:

$$\Phi_{ij} = a_{2j} - a_{2,j+1} + b_0 X_i + d_i(Z_{ij} - Z_{i,j+1}) \quad [5.11]$$

上式适用于公式 2.2 和公式 5.5,并且,

$$Z_{ij} = 1 \ (i = j) \qquad [5.12]$$
$$Z_{ij} = 0 \ (i \neq j)$$

对这一模型的解释与统一关联模型几乎相同。统一关联参数(b_0)表示的是 Φ_{1j} 到 Φ_{Rj} 连接线的斜率,它测量的是社会出身变量变化一个单位时,地位获得 j 相对于地位获得 j+1 的发生比增加多少个单位。然而,像行/列效应模型一样,并非所有的期望 logit 值都位于 Φ_{ij} 和 Φ_{Rj} 之间的连接线上。对角线效应参数使 logit 值在对角线附近上下移动(具体取决于 d_i 的符号以及 $i = j$ 或 $i = j+1$)。

准行效应模型、准列效应模型、准模型Ⅰ、准模型Ⅱ以及

准同质行与列效应模型可以用如下相似的表达式来定义：

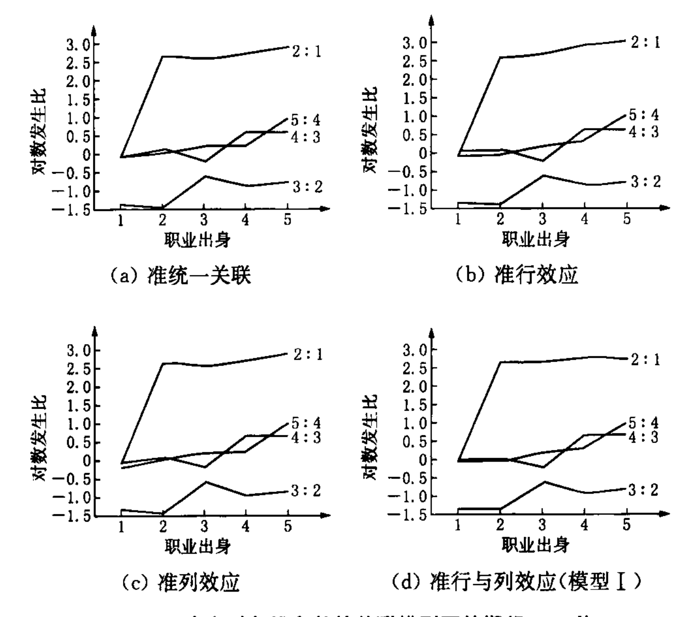
$$\Phi_{ij} = a_{2j} - a_{2,j+1} + b_0 X_i + b_{1i} + d_i(Z_{ij} - Z_{i,j+1}) \quad [5.13]$$

$$\Phi_{ij} = a_{2j} - a_{2,j+1} + (b_0 + b_{2j})X_i + d_i(Z_{ij} - Z_{i,j+1}) \quad [5.14]$$

$$\Phi_{ij} = a_{2j} - a_{2,j+1} + (b_0 + b_{2j})X_i + b_{1i} + d_i(Z_{ij} - Z_{i,j+1}) \quad [5.15]$$

$$\Phi_{ij} = a_{2j} - a_{2,j+1} + bu_i(v_j X v_{j+1}) + d_i(Z_{ij} - Z_{i,j+1}) \quad [5.16]$$

以上公式都适用于公式 2.2、公式 5.5 和公式 5.12。在这些模型中，行效应参数(b_{1i})和对角线效应参数(d_i)都会使 logit 值偏离从 Φ_{i1} 到 Φ_{iR} 的连接线。列效应参数(b_{2j})则使这些连接线的斜率发生变化，正如在列效应模型和模型 I 中那样。图 5.3 给出了准统一关联模型、准行效应模型、准列效应模



(a) 准统一关联 (b) 准行效应

(c) 准列效应 (d) 准行与列效应(模型 I)

图 5.3 含有对角线参数的关联模型下的期望 logit 值

型以及准模型Ⅰ在拟合美国流动表数据结果上的差别。

以删除对角线单元格后再建模这种思路为指导，可以得到许多不同的模型，并不仅仅包括上面所列举的这些。一些扩展形式的模型会在第6章中加以介绍。此外，也可以对上述一般化模型中的某些参数进行约束以得到新的模型。例如，公式5.15中的参数 d_i 可以被约束为一个常数 d 以适用所有的对角线单元格；对公式5.15的另一个约束是令两个行效应相等，例如，$b_{12} = b_{14}$。限于篇幅，本书无法列出所有可能的约束形式，但参数约束这种方法在第6章将要介绍的模型的最新发展中会显得尤为重要。

将上述模型应用于美国流动表数据后的拟合优度统计值见表5.3。模型的参数估计值见表5.4。拟合对角线单元格上的参数后，与表5.2中的相应模型相比较，大幅提升了模型的拟合优度。尽管所有这些模型在常规的显著性水平上都被拒绝了，但(屏蔽对角线单元格后的)行效应模型与模型Ⅰ的拟合优度比前几节中的任何一个模型都好。列效应则看似并不重要。缺乏模型Ⅱ的估计结果，就无法对这一结论的正确性作出严格的判断，但支持性的证据是模型Ⅰ在行效应模型的基础上并没有显著的改进。因此，对美国流动表数据来说，准行效应模型看似是一个最好的模型。正如图5.3(b)所显示的那样，较强的对角线效应冲淡了行效应。这些对角线效应使获得某一职业地位相对于获得低一级职业地位的发生比在每一类社会出身发生地位继承的地方出现了峰值。然而，社会出身对地位获得所具有的正向主效应依然是存在的。

表 5.3　父亲的职业与儿子的第一份职业之间的流动：各种准关联模型的估计结果

职业出身	职业获得 上层非体力	下层非体力	上层体力	下层体力	农民
(1) 准统一关联模型的期望频数					
上层非体力	1414	529.70	299.97	641.01	35.33
下层非体力	638.97	524	297.39	744.57	48.08
上层体力	841.64	691.71	856	1577.33	119.33
下层体力	861.90	829.93	755.90	3325	230.27
农民	344.50	388.66	414.75	1670.10	1832
(2) 准模型 I 的期望频数					
上层非体力	1414	533.87	278.30	651.80	42.03
下层非体力	716.37	524	271.66	698.22	42.76
上层体力	790.30	661.74	856	1665.56	112.39
下层体力	794.30	835.04	812.84	3325	235.82
农民	386.03	409.35	405.21	1617.41	1832
(3) 准同质模型 I 的期望频数					
上层非体力	1414	570.64	282.41	612.53	40.41
下层非体力	678.75	524	276.66	724.94	48.66
上层体力	791.25	651.67	856	1669.72	117.36
下层体力	825.97	821.84	803.62	3325	226.57
农民	391.03	395.84	405.31	1625.82	1832
	L^2	X^2	df	p	Δ
准完全流动模型	683.34	720.74	11	<0.01	0.055
准统一关联模型	73.01	73.45	10	<0.01	0.020
准行效应模型	34.91	35.08	7	<0.01	0.011
准列效应模型	66.69	67.81	7	<0.01	0.018
准模型 I	27.75	27.77	4	<0.01	0.010
准模型 II	24.11	23.78	4	<0.01	0.013
准同质模型 I	39.51	39.57	7	<0.01	0.013
准同质模型 II	39.28	39.24	7	<0.01	0.013

表 5.4 父亲的职业与儿子的第一份职业之间的流动：各种准关联模型的参数估计值（括号内的数字为标准误）

参数	准统一关联模型	准行效应模型	准列效应模型	准行与列效应模型			准同质行与列效应模型	
				Ⅰ	Ⅱ(a)	Ⅱ(b)	Ⅰ	Ⅱ(b)
统一关联效应	0.158* (0.006)	0.141* (0.008)	0.165* (0.014)	0.109* (0.009)	1.621	0.158	0.118* (0.015)	0.158
行效应								
1	—	0	—	0	0.680	5	0	5
2	—	0.061* (0.030)	—	0.040 (0.044)	0.325	3.863	−0.035 (0.038)	3.513
3	—	−0.053* (0.022)	—	−0.068* (0.028)	−0.083	2.560	−0.100* (0.025)	2.401
4	—	−0.100* (0.027)	—	0.144* (0.035)	−0.470	1.320	−0.136* (0.033)	1.313
5	—	0	—	0	−0.452	1.379	0	1.193
列效应								
1	—	—	0	0	0.748	5	0	5
2	—	—	−0.036 (0.027)	−0.044 (0.040)	0.207	3.269	−0.035 (0.038)	3.513
3	—	—	−0.048 (0.033)	−0.095* (0.038)	−0.070	2.383	−0.100* (0.025)	2.401

续表

参数	准统一关联模型	准行效应模型	准列效应模型	准行与列效应模型			准同质行与列效应模型	
				Ⅰ	Ⅱ(a)	Ⅱ(b)	Ⅰ	Ⅱ(b)
4	—	—	−0.002 (0.041)	−0.120* (0.050)	−0.410	1.297	−0.136* (0.033)	1.313
5	—	—	0	0	−0.475	1.089	0	1.193
对角线								
1	0.469* (0.054)	0.503* (0.081)	0.378* (0.072)	0.424* (0.158)	0.086	0.086	0.308* (0.155)	0.060
2	0.156* (0.055)	0.108 (0.058)	0.188* (0.061)	0.125 (0.082)	0.275	0.275	0.135 (0.164)	0.246
3	0.464* (0.047)	0.490* (0.048)	0.482* (0.048)	0.520* (0.050)	0.527	0.527	0.812* (0.098)	0.518
4	0.247* (0.035)	0.100 (0.055)	0.256* (0.036)	0.014 (0.070)	−0.000	−0.000	0.049 (0.067)	0.043
5	2.357* (0.061)	2.573* (0.079)	2.358* (0.110)	2.911* (0.164)	2.785	2.785	2.910* (0.140)	2.763

注：对模型Ⅱ无法计算标准误。* 表示 $p < 0.05$。

第6节 | 不同模型之间的参数相似性

留意一下表5.4中准行效应模型下对角线参数的估计值。d_3 与 d_1 的值相差不远，并且都显著地大于 d_2 或 d_4。而回顾前文，d_3 在边角模型和对角线模型中都比较小，在费瑟曼—豪斯模型中甚至不显著。这种反差是由定义参数 d_3 时所对照的基准不同造成的。在边角模型、对角线模型和费瑟曼—豪斯模型中，基准是准完全流动模型。对角线效应的测量是相对于其他交互效应存在时，调整后的边缘效应而言的(Pontinen, 1981)。但在(准)行效应模型中，基准是描述社会出身对地位获得的影响作用的曲线。该曲线反映了交互效应，边缘效应只是改变了曲线的截距。尽管如此，这四个模型下的期望频数却甚少变化。其中三个模型在第三个对角线单元格上所得到的期望频数完全相同，即使是在费瑟曼—豪斯模型这一例外情况下，误差也不过小于10%。

以边缘效应还是交互效应为基准，其本质差别在对四个模型进行实际解释时体现得十分清楚。边角模型、对角线模型和费瑟曼—豪斯模型表明，对于非出身于下层体力职业阶层和农民职业阶层的人来说，他们获得下层体力职业的情形相对较少，更多地是获得上层体力职业。并且，获得非体力职业的情形对非体力职业阶层出身的人来说特别普遍。获

得上层体力职业的发生比在上层体力职业阶层出身的人中最高，因为在特殊参数的作用下，其他阶层出身的人都被推向较高或较低的职业地位。根据行效应模型，获得较高职业地位的发生比随社会出身类别的上升而呈曲线模式增加。当应用于美国流动表数据时就意味着，当社会出身类别从下层体力职业上升到上层体力职业时，获得上层体力职业相对于获得下层体力职业的发生比只增加了一个很小的值。准行效应模型通过引入参数 d_i 改变了这一曲线模式。d_3 的估计值之所以比较大，是因为上层体力职业阶层出身的人获得上层体力职业的发生比实际上比行效应本身所预期的要大得多。

上面两段文字体现了两点：第一，要正确地解释某个参数，研究者就必须理解参数计算所对照的基准是什么。第二，不能指望从模型的期望频数来洞悉社会流动的过程与结构。任何模型的实质含义都在于参数的设定说明和参数的估计值。这一点使我们又回到了对等价形式的拓扑模型的讨论（见第 4 章有关拓扑模型的不确定性那一节）。下面的例子对前述讨论进行了更详细的说明，这其实是在邓肯(1978)的说明（Duncan & Schuman，1980）上所做的某种改动。

第7节 | 再论等价模型

试想一下 3 × 3 流动表的两个模型。第一个模型是 QPM;第二个模型包括一个统一关联参数(b)、一个同质行—列参数(c)和一个适用于全部三个对角线单元格的参数(d),这里姑且称之为“对角线行—列模型”(DRC)。这两个模型的期望频数分别为:

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + d_i Z_{ij} \qquad [5.17]$$

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + bX_iX_j + c(X_i + X_j) + dZ_{ij} \qquad [5.18]$$

以上两式都适用于公式 2.2、公式 5.5 和公式 5.12。两个模型都有三个参数和一个自由度。更重要的是,两个模型的期望频数是完全一样的(Duncan, 1978)。正因为如此,其中一个模型的参数可以表达为另一个模型的参数的函数:

QPM 为 DRC 的函数	DRC 为 QPM 的函数
$d_1 = d + 2(b - c)$	$b = (d_1 - 2d_2 + d_3)/6$
$d_2 = d - b$	$c = (d_3 - d_1)/4$
$d_3 = d + 2(b + c)$	$d = (d_1 + 4d_2 + d_3)/6$

[5.19]

这两个模型只在统计意义上具有等价性。就实际意义而言,它们是不同的。QPM 假定不同职业类别之间在继承性强度

上的不同，以及在移动者中社会出身和地位获得之间的独立性。DRC则假定所有职业类别的继承性强度是相同的，并假定社会出身对获得较高职业地位的发生比的影响呈曲线模式。为区分这两个模型，必须借助于第4章讨论过的一系列原则，包括对称性、简约性和可重复性等。但在这里，对称性原则和简约性原则发挥不了作用，因为两个模型都具有对称性结构，也都有三个交互项参数，所以只能依赖于可重复性这一原则。

第8节 | 英国和丹麦的流动比较

学者经常分析的英国和丹麦的 3 × 3 流动表，在这里可以被用来体现何为可重复性。数据直接来自古德曼(1965：572)的著作，但原始数据请见格拉斯(1954)和斯瓦拉斯托加(Svalastoga，1959)的著作。职业类别被简单地划分为上层、中层和下层，具体的编码细则见古德曼(1965)的著作。这两个流动表被视为一个三维列联表(三维指的是社会出身、地位获得、国家)。克洛格(1982a)的方法被用于对 DRC 模型中的参数 b、c、d 进行跨国家约束，古德曼(1969b)的方法被用于对 QPM 的参数进行跨国家约束。

分析集中于 DRC 模型。最开始是对任何一个模型都不进行跨国家约束。随后，DRC 的参数 b 被约束为在英国和丹麦两个国家之间相等。最后，再依次分别将参数 c 和参数 d 约束为在两个国家之间相等。这四个模型依次被命名为模型 U、模型 B、模型 C、模型 D。公式 5. 19 表达的 DRC 各参数之间的关系意味着，如果模型 D 被接受为最优模型，则参数 d_i 在国家之间一定相等。但是，如果模型 B 或模型 C 被接受为最优模型，则公式 5. 19 并不能表明跨国家约束是成立还是不成立。事实上，如果模型 B 或模型 C 为真，则英国流动模型中的 d_i 值就不可能等于丹麦流动模型中的 d_i 值。因

此，如果模型B或模型C被选中为DRC模型的具体形式，它一定比相应的QPM模型的具体形式更简约，DRC与QPM之间孰优孰劣就一目了然。运用同样的逻辑，如果三个d_i参数中只有一个或两个可以被约束为在国家之间相等，那么，QPM就比DRC更简约。

上述模型的比较结果见表5.5。正如古德曼在对两国数据所作出的最初分析中指出的，QPM对数据的拟合非常好，但不可避免地，DRC也同样如此。将参数b、c、d约束为在英国和丹麦两国之间相等后，模型的拟合优度并没有显著下

表5.5 流动表的跨国比较结果

	模型U		模型B		模型C		模型D	
	英国	丹麦	英国	丹麦	英国	丹麦	英国	丹麦
(1) 拟合优度								
L^2	1.44		1.67		2.19		4.11	
X^2	1.44		1.67		2.19		4.10	
df	2		3		4		5	
Δ	0.005		0.005		0.006		0.010	
(2) DRC模型的参数								
b	0.563	0.599	0.578	0.578	0.578	0.578	0.581	0.581
c	−0.099	−0.134	−0.099	−0.135	−0.113	−0.113	−0.115	−0.115
d	0.175	0.218	0.164	0.235	0.164	0.240	0.191	0.191
(3) QPM模型的参数								
d1	1.501	1.684	1.519	1.660	1.545	1.621	1.583	1.583
d2	−0.388	−0.380	−0.413	−0.343	−0.414	−0.338	−0.390	−0.390
d3	1.103	1.147	1.122	1.122	1.094	1.169	1.122	1.122
(4) 对数优比的期望值								
(θ_{11})	1.113	1.304	1.105	1.317	1.131	1.283	1.193	1.193
(θ_{12})	0.388	0.380	0.413	0.343	0.414	0.338	0.390	0.390
(θ_{21})	0.388	0.380	0.413	0.343	0.414	0.338	0.390	0.390
(θ_{22})	0.715	0.766	0.709	0.779	0.680	0.832	0.732	0.732

降。这三个约束条件加在一起，只不过使 L^2 值增加了 2.67($df = 3$, $p > 0.5$)。这种结果并不能使我们在 QPM 和 DRC 之间作出选择。于是就存在两种结论：其一，存在一个大体上陡峭但在某些地方有所缓和的斜率($b = 0.581$)。该斜率指的是，社会出身每上升一个类别时，获得较高职业地位的发生比的增加值，"某些地方"是指当社会出身类别从下层上升为中层时，存在一个对所有的职业类别而言强度都适中的继承性。其二，存在一种准完全流动模式，它在职业序列的两端有较高的继承性，但在中间则具有较高的"非继承性"。导致非继承性现象的原因有很多，但我认为，这一过程并不容易发生。此外，它在更大的流动表中也并没有出现。就英国和丹麦这两个国家的流动表数据而言，我更倾向于 DRC 模型，因为它的参数值更合理。

邓肯(1979:801)的结论用在这里甚为恰当：

> 模型(4)(即准统一关联模型)对比较研究特别具有吸引力，为进行这种比较，它会非常严格地检验一个假设，即反映行与列交互效应的唯一的模型参数在两个及两个以上的研究总体之间是相等的，或者假设该参数随时间和研究的总体变化符合研究者所期望和指定的某种数学函数。

本章已经显示了各种关联模型在概括和比较不同社会流动结构上的能力。

第6章

新的发展

前面所介绍的模型都可直接应用于流动表数据，但近来最有意思的研究是对这些基本模型作出进一步的发挥。本章将回顾克洛格(1981)、布里格(1981)和豪特最近的论文。由于每篇论文都展示了一个独特的模型，所以它们不可能被一起讨论。但它们都探讨的几个共同议题值得先行说明，以免这些普遍性的要点被论文的细节所淹没。山口(Yamaguchi, 1983)和洛根(Logan, 1983)也讨论了许多这样的议题，但限于篇幅，在此无法对他们的论文进行详细介绍。

第一个议题是关于适合的汇总层次。布里格、山口和豪特分析了来自OCG的17×17流动表数据。其他人分析的是较小的流动表。在汇总层次这个议题之下就是职业类别的合并问题。布里格和古德曼(1981)讨论了合并时的准则。

第二个议题是如何处理对角线。大多数研究者都沿袭已经为大家所接受的做法，即删除对角线单元格或者对它们进行精确的拟合(两者是等同的)。克洛格通过区分潜在的移动者和滞留者群体来说明对角线上的代际地位持续性。豪特通过引入独立变量来明确地为对角线进行建模，以说明代际继承的模式。

第三个议题是在分析中引入独立变量。山口、洛根和豪

特对此都有论述。山口和豪特比较了亚群体之间的流动模式;洛根将独立变量对流动结构的影响用公式进行了正式的描述。

第四个议题涉及职业类别的度量。大多数研究者都使用如公式 5.5 那样的统一间隔尺度,或者只是它的某种变形。豪特使用邓肯(1961)的 SEI 均值来对职业类别进行赋值。

第五,豪特和山口提出了其他维度(而非地位维度)流动的可能性。

最后,布里格和克洛格考虑了流动的阶级隔阂议题。尽管他们的研究路径相当不同,但都致力于寻找潜在的社会阶级结构。布里格将阶级定义为某些职业类别的集合,但克洛格视阶级为一个嵌入的维度(它根据职业的本质区别进行划分),他对阶级的标准分类体系更感兴趣。这六个议题构成了最近的流动研究文献的主题。

第 1 节 | 潜在的流动结构

与本章所要讨论的其他模型相比,克洛格的模型显然与前面所介绍的基本模型关系不大。但正如克洛格自己指出的,QPM、边角模型、FH5 以及其他几个基本模型都可以用潜在结构模型来表达。具体而言,这又包括一个标准潜在结构模型(Lazarsfeld & Henry, 1968; Goodman, 1974; Haberman, 1979:541—552)和一个全新的准潜在结构模型。潜在结构分析的基本观点是,一个或多个未被观测到的变量解释了已被观测到的变量之间的相关关系。在社会流动分析的情境下,它等于是说社会出身与地位获得之间其实是相互独立的,只不过它们都与某个未被观测到的变量存在相关性而已。克洛格对于发现一个或多个潜在变量来解释英国和丹麦流动表数据中社会出身与地位获得之间的关系十分感兴趣。在本节中,我将美国流动表数据也加入进来进行比较。

我用 π_{ij} 表示流动表中第(i, j)格的期望频数比例,即 $\pi_{ij} = F_{ij}/N$,其中,F_{ij} 由某个模型所决定(模型的具体形式尚不确定)。用 t 表示未被观测到的第三个变量(Z)的分类,π_{ijt}^{XYZ} 表示在未被观测到的三维列联表中,第(i, j, t)单元格的期望频数比例。那么,观测频数比例(π_{ij})就等于 Z 取各个类别

时的 π_{ijt}^{XYZ} 相加之和：

$$\pi_{ij} = \sum_{t} \pi_{ijt}^{XYZ} \qquad [6.1]$$

基本的潜在结构模型可以用潜在变量的边缘分布(π_{t}^{Z})以及当一个随机潜在类别 t 对应社会出身类别 i(π_{it}^{XZ})和地位获得类别 j(π_{jt}^{YZ})时的条件概率来表达：

$$\pi_{ijt}^{XYZ'} = \pi_{t}^{Z}\pi_{it}^{XZ}\pi_{jt}^{YZ} \qquad [6.2]$$

将公式 6.2 代入公式 6.1，就得到了“潜在结构分析的基本方程式”(Clogg，1981:839)：

$$\pi_{ij} = \sum \pi_{t}^{Z}\pi_{it}^{XZ}\pi_{jt}^{YZ} \qquad [6.3]$$

请留意，该模型并没有像其他模型那样，将边缘效应与交互效应进行区分。观测到的边缘效应与交互效应都被看做只是反映了潜在的分布结构与交互关系。研究总体被假设为沿着各个潜在类别分布(在本例中有五个潜在类别)。假定某一个体属于某一潜在类别，则在五个可能的社会出身类别[①]与五个可能的地位获得类别中任取一种组合的概率是固定的。观测到的交互模型只不过反映了潜在的边缘分布与条件概率而已。注意，这正是潜在结构模型的一个问题所在：尽管可以通过拟合观测到的边缘分布来确定潜在类别，但模型的参数并不一定将边缘效应与交互效应进行区分。观测到的边缘分布是由两个原因引起的：一个是潜在的边缘分布，一个是潜在变量与观测变量之间的交互关系。

克洛格为 5×5 英国/丹麦流动表和 8×8 英国流动表设

① 原文为“四个可能的社会出身类别”，为笔误，已改正。——译者注

想了许多模型。他为 5×5 流动表选中的模型包含五个潜在类别。前两个类别是潜在的上层阶级与潜在的下层阶级。前者通过当 $i=5$ 和 $t=1$ 时，约束 $\pi_{it}^{XZ}=0$ 来定义，即将社会出身类别最低的人从潜在类别 1 中排除出去；后者通过当 $i=1$[①] 和 $t=2$ 时，约束 $\pi_{it}^{XZ}=0$ 来定义，即将社会出身类别最高的人从潜在类别 2 中排除出去。其他三个潜在类别分别由职业类别 1、职业类别 3 和职业类别 5 中的潜在"滞留者"构成。它们分别通过当 $i=j=1$ 和 $t=3$ 时约束 $\pi_{it}^{XZ}=\pi_{jt}^{YZ}=1$，当 $i=j=3$ 和 $t=4$ 时约束 $\pi_{it}^{XZ}=\pi_{jt}^{YZ}=1$，当 $i=j=5$ 和 $t=5$ 时约束 $\pi_{it}^{XZ}=\pi_{jt}^{YZ}=1$ 来定义。

该模型的结果见表 6.1。从中可见，它大致拟合了英国流动表数据，很好地拟合了丹麦流动表数据，但却不能像其他模型那样较好地拟合美国流动表数据。成为潜在的上层阶级成员的发生比在英国为 0.318(0.220/0.691 = 0.318)，比在丹麦的 (0.251/0.624 = 0.402) 大约低了 20%，比在美国的 (0.296/0.567 = 0.522) 大约低了 40%。上层阶级成员具有上层非体力职业出身或获得上层非体力职业地位的条件概率在美国比在其他两个国家高出许多。我们尚不清楚这些差别到底反映的是三个社会在分层过程上的实质性不同，或只是时期上的不同(在两个欧洲国家开展的研究比在美国进行的研究要早 20 多年)，又或者只是编码上的不同。不管是哪一种原因，条件概率的这种差异都反映了实际观察到的边缘分布的差异。但至于它是否也反映了交互效应模式的差异，则不得而知了。下层阶级的条件分布在国家之间

---

① 原文为"$j=1$"，为笔误，已改正。——译者注

的差别并不大。

表6.1 父亲的职业与儿子的现职(英国和丹麦)或第一份职业(美国)之间的流动:包括五个潜在类别的模型结果

	人数比例(给定潜类 t)	社会出身类别为 i 的条件概率(给定潜类 t)		职业获得类别为 j 的条件概率(给定潜类 t)	
取值	(π_t^Z)	(π_{i1}^{XZ})	(π_{i2}^{XZ})	(π_{j1}^{YZ})	(π_{j2}^{YZ})
(1) 英国					
1	0.220	0.111	0	0.077	0.000
2	0.691	0.509	0.043	0.438	0.063
3	0.012	0.178	0.137	0.165	0.117
4	0.014	0.202	0.560	0.254	0.510
5	0.062	0	0.260	0.067	0.310
(2) 丹麦					
1	0.251	0.071	0	0.084	0.010
2	0.624	0.450	0.032	0.383	0.022
3	0.006	0.342	0.242	0.375	0.195
4	0.059	0.137	0.466	0.125	0.505
5	0.059	0	0.260	0.032	0.269
(3) 美国					
1	0.296	0.378	0	0.342	0.129
2	0.567	0.348	0.018	0.233	0.141
3	0.031	0.249	0.199	0.122	0.136
4	0.019	0.015	0.524	0.285	0.556
5	0.086	0	0.260	0.018	0.039

注:取值表示 π_t^Z 中 t 的取值、π_{i1}^{XZ} 和 π_{i2}^{XZ} 中 i 的取值以及 π_{j1}^{YZ} 和 π_{j2}^{YZ} 中 j 的取值。

第2节 从流动模式推论阶级结构

布里格(1981)提出了一个模型,不仅可以将一个大的(17×17)流动表的职业分类经过合并得到数量较少的类别集合,还可以同时对这些类别集合之间的交互参数进行估计。借用韦伯对阶级的定义,即阶级表示"其内部的个体或代际流动较易发生并且具有某种典型性"(Breiger, 1981: 579),布里格将社会阶级定义为某些职业类别的集合,在某一阶级所对应的职业流动分表中,职业获得与职业出身之间必须是相互独立的。这一模型的建立步骤如下:(1)将流动表中的R个职业类别划割为C个阶级。将这种划割方法同时应用于R×R表的行与列,得到C^2个分表。(2)对不包含对角线单元格的分表拟合完全流动模型;对包含对角线单元格的分表拟合准完全流动模型。将每一个分表模型的L^2和自由度相加,以检验职业出身与职业获得在"阶级"内部相互独立这一零假设。(3)如果不能拒绝上述零假设,就将每一个分表中的频数相加,得到一个C×C表。然后再用第5章所介绍的模型来拟合这个C×C表。

布里格尝试了相关研究文献所提议的好几种划割方法,但最终只选择了一个八阶级划割法,并将其应用于从OCG-II得到的父代职业与子代第一份职业之间的17×17流动表

(该流动表是表1.1的一个细化版本)。这八个阶级对图1.1中17个职业类别的合并方式如下:(1),(2,3,4),(5),(6,7),(9,10),(8,13,14),(11,12,15),(16,17)。

为检验这一划割方法是否适当,我们根据步骤(2)来分别拟合独立模型和准完全流动模型。八阶级划割法本来可能产生64个检验,但由于布里格令其中的几个阶级只包含一个或两个职业类别,所以造成了31个检验的自由度为0。这些无法检验的分表包括了30.2%的样本量。剩余的33个检验的L^2值和自由度见表6.2。根据布里格(1981:595)的报告,这些L^2值相加之和为76.9,自由度相加之和为69。拟合结果是好的($p > 0.20$)。隐含的187个参数解释了17×17流动表中大部分的相关关系。

表6.2 父亲的职业与儿子的第一份职业之间的流动:布里格模型中33个检验的L^2值

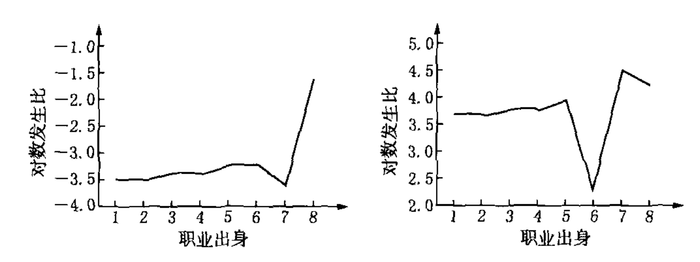
	布里格的阶级分类							
阶级	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ	Ⅶ	Ⅷ
Ⅰ	—	—	—	—	—	—	—	—
Ⅱ	—	1.88* (1)	—	2.02 (2)	1.80 (4)	5.33 (2)	8.94 (4)	1.86 (1)
Ⅲ	—	—	—	—	—	—	—	—
Ⅳ	—	4.78 (2)	—	—	0.55 (2)	0.05 (1)	0.05 (2)	1.40 (1)
Ⅴ	—	0.99 (4)	—	1.98 (2)	0.04 (1)	1.09 (2)	0.62 (4)	8.68** (2)
Ⅵ	—	0.47 (2)	—	0.90 (1)	0.22 (2)	—	8.74** (2)	0.70 (1)
Ⅶ	—	6.18 (4)	—	2.47 (2)	4.21 (4)	1.79 (2)	0.27 (1)	0.43 (2)
Ⅷ	—	0.94 (2)	—	0.32 (1)	0.22 (2)	3.13 (1)	1.56 (2)	—

注:*表示括号中的数字是自由度。**表示$p < 0.05$。

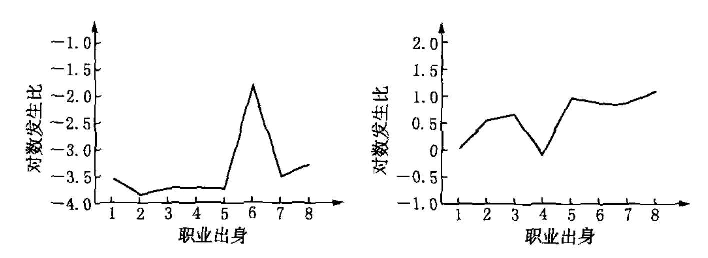
表 6.3 给出了步骤(3)的结果——分析职业类别被重新划割后所得到的流动表中的相关关系。邓肯(1979)和古德曼(1979a)所考虑过的模型没有一个可以拟合数据。布里格对准行效应模型进行了改动，具体方法是屏蔽农民职业阶层这一行上的三个单元格，即那些出身于农民职业阶层但成为阶级Ⅳ、阶级Ⅴ和阶级Ⅵ成员的人，结果对数据的拟合相当好。图 6.1 展示了该模型的期望发生比。对角线效应占据了主导性的地位，在非对角线区域，几乎见不到社会出身对地位获得的影响。

有证据表明，布里格的模型可能在很大程度上掩盖了非对角线区域所存在的重要交互关系，即美国流动表数据中父亲的职业对儿子第一份职业的影响。其中一些证据涉及(职业的)地位、自主性和培训模型(SAT)的应用，这将在下文中予以介绍。另外一些证据则涉及布里格在将单元格进行合并以生成新的“阶级流动表”时所采用的准则是否一定正确。别忘了，在用来检验合并方法是否恰当的分表中，有将近一半因为自由度为 0 而没有派上用场。因此，古德曼(1981)提出了合并职业类别的准则，以避免发生分表的自由度为 0 这种情形。根据他的方法，如果研究者想检验职业类别 2 与职业类别 3 是否可以合并，就需要三个分表。第一个分表是由第二行和第三行构成的 $2 \times (R-2)$ 表，第二列和第三列被从这个表中删除。第二个分表是由第二列和第三列构成的 $(R-2) \times 2$ 表，第二行和第三行被从这个表中删除。第三个分表是一个不完整的 3×3 表：

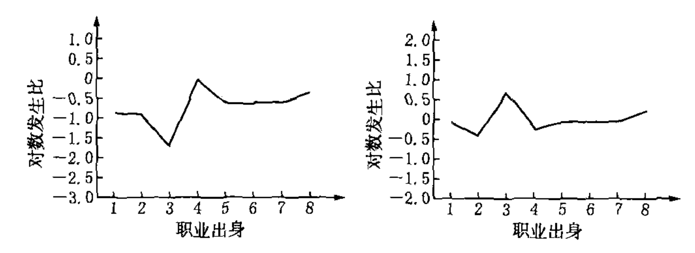
—	f_{23}	$n_{2.} - f_{22} - f_{23}$
f_{32}	—	$n_{3.} - f_{32} - f_{33}$
$n_{.2} - f_{22} - f_{32}$	$n_{.3} - f_{23} - f_{33}$	—



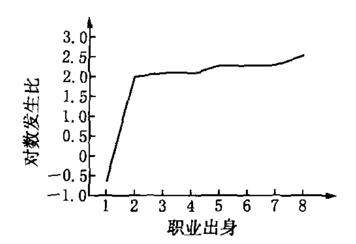
(a) 有产的专业人士：受雇的中产阶级 (b) 受雇的中产阶级：小业主



(c) 小业主：程式化的白领工人 (d) 程式化的白领工人：半自主的技术工人



(e) 半自主的技术工人：制造业工人 (f) 制造业工人：传统的工人阶级



(g) 传统的工人阶级：农业阶级

图 6.1 布里格模型下的期望 logit 值

表 6.3　父亲的职业与儿子的第一份职业之间的流动：布里格的分析结果

阶级出身	阶级获得 Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ	Ⅶ	Ⅷ
（1）M6 模型的期望频数：								
Ⅰ	25	129.02	1.85	48.94	15.86	23.52	19.26	1.56
Ⅱ	31.66	1222	13.85	460.04	186.95	347.88	357.32	36.29
Ⅲ	12.35	311.44	33	199.59	84.04	162.02	172.42	18.14
Ⅳ	14.75	378.22	7.16	291	107.33	210.43	227.73	24.37
Ⅴ	15.39	469.42	10.58	440.29	490	523.11	673.47	85.74
Ⅵ	18.87	566.74	12.58	515.38	258.50	1464	752.59	94.34
Ⅶ	14.00	479.22	12.12	565.76	323.32	846.04	1283	174.54
Ⅷ	11.98	399.94	9.87	344	328	822	902.22	1832

（2）参数估计值

	统一关联模型	行效应模型	列效应模型	行与列效应模型Ⅰ 行效应	行与列效应模型Ⅰ 列效应	M6 模型
统一效应	0.067	0.082	0.058	0.071		0.078
Ⅱ	—	−0.139	0.063	−0.109	0.044	−0.149

续表

	统一关联模型	行效应模型	列效应模型	行与列效应模型 I		M6 模型
				行效应	列效应	
Ⅲ	—	−0.098	0.044	−0.089	0.039	−0.107
Ⅳ	—	−0.034	0.026	−0.025	0.022	−0.046
Ⅴ	—	−0.125	−0.021	−0.116	−0.026	−0.142
Ⅵ	—	−0.032	−0.032	−0.027	−0.040	−0.049
Ⅶ	—	−0.064	0.028	−0.049	0.013	−0.102
		L^2	X^2	df	p	Δ
统一关联模型		167.32	166.83	40	<0.01	0.030
行效应模型		99.81	97.86	34	<0.01	0.020
列效应模型		124.43	123.92	34	<0.01	0.024
行与列效应模型 I		67.95	66.29	28	<0.01	0.015
M6 模型		39.29	37.80	31	0.19	0.009

注:(1)栏中的整数表示该单元格被精确地拟合。

其中，n_i 表示行边缘和，n_j 表示列边缘和。如果完全流动模型能够拟合前两个分表，并且准完全流动模型能够拟合第三个分表，那么，职业类别 2 与职业类别 3 就可以合并。这三个检验的自由度分别为(R－3)、(R－3)、1。它们在表6.4中被称为“行”、“列”和“对角线”检验(表 6.4 给出了布里格模型中合并的所有职业类别配对的检验结果)。由于这三个检验是相互独立的，在同一“阶级”内部社会出身与地位获得之间不存在相关关系这一零假设下(被合并在一起的职业类别就构成了一个“阶级”)，它们相加所得就服从自由度为 2(R－3)＋1 的卡方分布。使用古德曼的职业类别合并准则所得出的结果与布里格的结果大相径庭。没有任何一对职业类别能够全部通过上述检验。关于准完全流动模型的拟合性检验，只有两对职业类别没有通过。可以说在很大程度上，非对角线区域的相关性信息在合并过程中丢失了。

为何使用古德曼的合并准则得出的结果与布里格的结果存在如此根本性的差异？哪里有矛盾呢？其实没有矛盾。古德曼的合并准则要求，在对行内的职业类别进行合并检验时，具有一种社会出身(i)相对于具有另外一种社会出身(i′)的发生比对所有的地位获得类别(j)来说，必须都是一个常数(除了当 $j = i$ 和 $j = i'$ 时)；在对列内的职业类别进行合并检验时，获得一种职业地位(j)相对于获得另外一种职业地位(j′)的发生比对所有的社会出身类别(i)来说，必须都是一个常数(除了当 $i = j$ 和 $i = j'$ 时)。但布里格的准则允许这些发生比存在较大的变化，只要在阶级内部具有恒定性即可。虽然发生比在阶级内部没有什么不同，但在阶级之间却具有很大的差异。在根据步骤(2)来合并职业类别时，布里格掩盖了一部分相关性，而正是它造成了阶级之间发生比的不同。

表 6.4 将古德曼的职业类别合并准则应用于布里格的阶级分类的结果

阶级分类	行 (df = 14)	列 (df = 14)	对角线 (df = 1)	合计 (df = 29)
阶级Ⅱ				
专业人士(受雇)＋经理人员	33.05*	19.77	0.08	52.90*
专业人士(受雇)＋销售人员(非零售业)	16.98	37.17*	0.05	54.20*
经理人员＋销售人员(非零售业)	7.63	5.94	4.06*	17.63
阶级Ⅳ				
文职人员＋销售人员(零售业)	13.22	37.80*	0.35	51.37*
阶级Ⅴ				
技术工人(其他行业)＋技术工人(建筑业)	26.86*	34.66*	0.25	61.77*
阶级Ⅵ				
技术工人(制造业)＋操作工人(制造业)	47.43*	29.24*	1.07	77.74*
技术工人(制造业)＋搬运工人(制造业)	78.07	52.39*	1.83	132.29*
操作工人(制造业)＋搬运工人(制造业)	27.78*	20.16	0.09	48.03*
阶级Ⅶ				
服务业人员＋操作工人(非制造业)	21.05	32.59*	1.76	55.40*
服务业人员＋搬运工人(非制造业)	42.96*	43.32*	2.85	89.13*
操作工人(非制造业)＋搬运工人(非制造业)	23.44*	24.56*	0.07	48.07*
阶级Ⅷ				
农场主或农场经营者＋农业耕作者	34.80*	32.36*	23.19*	90.35*

注：* 表示 $p < 0.05$。

第3节 职业流动中的地位、自主性和培训

豪特对流动表的建模方法与克洛格和布里格的方法十分不同。他的论文考虑了三个主题：邓肯(1979)和古德曼(1979a)所使用的得分尺度的随意性、用独立变量来解释继承性模型，以及职业的其他特征(而非地位特征)对大表(17×17)中职业流动结构的重要性。为方便起见，哈伯曼(1974)的线性乘线性交互模型(公式5.4)对职业分类采用了常见的单一尺度得分法(见公式5.5)。任何已知的得分都可以被代入。研究者考虑最多的职业特征是职业地位，因此，公式5.5被取代为：

$$X_i = S_i, \; X_j = S_j$$

其中，S_i 和 S_j 表示职业类别 i 和 j 的 SEI(Duncan, 1961)得分均值。还可以在模型中加入职业特征的其他维度，所使用的标记如下：A_i(如果从事职业类别 i 的人有一个直接上司或监管人，则将发生比的值乘以－1)、T_i(表示从事职业类别 i 的人受到过特定的职业培训)。以上两个变量都用综合社会调查1972年至1980年的合并样本来进行估计(NORC, 1980；详见 Hout，待出)。

此外，像公式5.13那样在模型中引入虚拟变量(Z_i)来加入对角线效应，它在对角线单元格上取1(否则取0)。当该虚拟变量乘以由尺度变量(地位、自主性或培训)所构成的函数时，这一效应的系数表示的是对角线单元格偏离表格其他区域的关联模式的情形。这一方法不同于用R个变量来精确拟合对角线单元格这种常用的"准"模型，它只用了三个变量来反映对角线效应。这三个变量的含义是，对于某个给定的职业类别，观测到的对角线频数偏离非对角线区域关联模式的情形是与该职业类别的地位、自主性和培训成比例的。

在这一模型下，期望频数服从下述公式：

$$\log(F_{ij}) = a_0 + a_{1i} + a_{2j} + b_1 S_i S_j + b_2 A_i A_j + d_1 Z_i S_{i2} + d_2 Z_i A_{i2} + d_3 Z_i T_i \qquad [6.4]$$

其中，当 $i = j$ 时，$d_i = 1$(否则就等于0)。注意，T_i 并没有被包括在主效应中，在对角线效应中也没有取平方。因为初步分析发现，主效应并不重要，并且 T_i 的线性形式比 T_i^2 的拟合度更好。模型中还有四项并没有在上述方程中体现出来。它们都是虚拟变量，反映了流入与流出农民职业阶层的情形比预期的少。这四个变量允许个体在农民职业阶层与另外四个职业阶层之间的流动模式存在差异，它们分别是：(1)上层非体力职业阶层；(2)下层非体力职业阶层；(3)上层体力职业阶层；(4)下层体力职业阶层。这些变量的效应被假定为是对称的，也就是说，不管是社会出身还是地位获得为农民职业阶层，所对应的变量效应是相同的。

在豪特的论文中，上述SAT模型被应用于父亲的职业与儿子现在的职业所构成的流动表。我将该模型分别应用

于1962年和1973年父亲的职业与儿子的第一份职业所构成的流动表，并且将白人样本和黑人样本分开，得到了表6.5中的结果。模型并不拟合白人样本的数据，尽管L^2与自由度的比值和其他许多模型相当。但模型却能够拟合黑人样本的数据。在解释这一反差时，必须小心谨慎。请留意，黑人样本数据与白人样本数据的Δ值相差不大。这种Δ值之间的少许差别表明，L^2之间的差异更有可能是由样本规模的差异，而非模型本身的拟合度差异所导致的。当对白人样本的几个出生组分别进行分析时，模型对每一个出生组数据的拟合度都是可以接受的。

表6.5 父亲的职业与儿子的第一份职业之间的流动(分种族和年份):SAT模型的估计结果

		1962年		1973年	
		白人	黑人	白人	黑人
(A) 主效应					
地位	(b1)	0.735*	0.430*	0.618*	0.642*
自主性	(b2)	0.450*	0.783*	0.354*	0.310*
(B) 对角线效应					
地位	(d1)	−0.202*	−0.250	−0.194*	−0.096
自主性	(d2)	0.142*	0.028	0.100*	0.155
培训	(d3)	0.221*	0.168*	0.201*	0.154*
(C) 拟合优度					
L^2		529.38	228.05	631.51	161.33
X^2		555.24	321.19	660.24	158.96
df		247	247	247	247
P		<0.05	>0.50	<0.05	>0.50
Δ		0.075	0.143	0.046	0.089
N		10022	836	18141	1833

注：*表示$p<0.05$。

模型的参数估计值几乎从各个方面(除了一个方面之

外)体现了预期的流动模式。职业的地位特征和自主性特征对总体流动性的影响很强,并且对黑人和白人来说都非常显著。自主性对继承性的影响在黑人和白人中都是正向的,但只有在白人中才显著。职业培训特征的影响在两个年代和两个人群中都很强,而且是正向的。出乎意料的结果是,职业的地位特征对继承性的影响是负向的。它之所以出现在公式中,唯一的原因是要用它来控制职业培训特征中所包含的地位特征成分。

与布里格的模型相比,豪特的模型揭示出,职业地位对代际流动的实质性影响更大,而且,它用职业之间在培训和自主性特征上的差异解释了大部分的对角线效应。相比之下,布里格的模型只是简单地删除了对角线单元格而已。最后,豪特的模型还简约很多,它仅仅拟合了九个交互项参数,而布里格的模型则拟合了 195 个。

第7章

流动表在家庭和宗教研究中的应用

流动表的行分类与列分类具有一致性，这种特性在其他研究领域中也可以见到。当行与列在分类上相同时，本书所介绍的许多模型都可加以应用。当这种分类还可以进行某种排序时，则前文中提到过的所有模型都可以应用。本章将回顾把流动表分析方法应用于其他研究领域的四个例子。

第1节｜丈夫与妻子的职业

在夫妻都有工作收入的家庭中，将丈夫的职业与妻子的职业进行交叉分类而得到的列联表看起来非常像一个流动表。不仅行与列在分类上具有一致性，而且这种分类也是职业分类。不同的是，流动表中的行与列表示的是人们在两个时间点上所从事的职业类别。但在夫妻职业列联表中，行与列则表示在同一时间点上一个人与另一个人所从事的职业类别。豪特(1982)分析了1978年3月时，丈夫职业与妻子的职业之间的相关性。表7.1给出了他的分析结果。

表7.1　双收入家庭中丈夫的职业与妻子的职业之间的相关性分析(1978年3月)

	妻子的职业				
丈夫的职业	上层非体力	下层非体力	蓝领	服务业	农业
(1) 观测频数					
上层非体力	1708	1872	236	416	26
下层非体力	405	836	142	192	3
蓝领	741	2242	1297	1125	34
服务业	149	279	112	245	3
农业	62	97	34	67	114
(2) 同质行与列效应模型I的期望频数					
上层非体力	1708	1877.55	230.34	422.03	20.08
下层非体力	398.61	836	145.31	192.86	5.22
蓝领	751.68	2233.67	1297	1121.68	34.97
服务业	135.44	291.53	110.30	245	5.73
农业	71.27	87.25	38.05	63.43	114

续表

	妻子的职业				
丈夫的职业	上层非体力	下层非体力	蓝领	服务业	农业
(3) 同质行与列效应模型 I 的参数估计值					
统一效应	−0.090*				
行/列效应		0.299*	−0.108*	−0.175*	
模型		L^2	X^2	df	p
完全流动模型		2263.38*	3921.35*	16	<0.05
准完全流动模型		364.14*	368.14*	11	<0.05
准统一关联模型		151.78*	163.07*	10	<0.05
准行效应模型		78.70*	85.68*	7	<0.05
准列效应模型		111.45*	110.96*	7	<0.05
准模型 Ⅰ		5.30	5.08	4	0.18
准模型 Ⅱ		14.73*	10.49*	4	<0.05
准同质模型 Ⅰ		9.78	9.49	7	0.20
准同质模型 Ⅱ		19.38*	15.56*	7	<0.05

注：*表示 $p<0.05$。

在流动表中，社会出身与地位获得之间存在的时间先后顺序使得因果变量的区分非常清楚。但在夫妻职业列联表中，并不存在这种表示因果关系的时间顺序。两个变量之间的影响是相互的，或者说，它们之间的相关性可能是由共同的外在力量所导致的，例如，基于教育和阶级的婚姻匹配。这种互为因果的特点使得在建模时需要优先考虑对称性。因此，就要特别留意具有对称性的相关模型：统一关联模型和模型Ⅰ与模型Ⅱ的同质形式。正如表 7.1 的最后一栏所示，对 1978 年的夫妻职业列联表数据来说，准同质行与列效应模型（模型Ⅰ）是最优模型。

丈夫的职业与妻子的职业之间的相关性大多体现在对角线上，完全流动模型与准完全流动模型的 L^2 值之差达到了 1899.24（自由度个数之差为 5）。其余的相关性体现在同质

行—列效应和一个负的统一关联参数上。在列效应与负的统一关联参数的共同作用下，妻子从事下层非体力职业相对于从事蓝领职业的发生比具有正的斜率，但妻子从事服务职业相对于从事农民职业的发生比以及从事上层非体力职业相对于从事下层非体力职业的发生比则具有负的斜率。然而，需要注意的是，上述斜率都排除了对角线效应（尤其是上层非体力职业阶层与农民职业阶层的对角线效应）的影响，因此，非对角线区域的相关关系的负斜率其实就相当于校正了在职业层级两端存在的极大的对角线效应。这在图7.1中表现得十分清楚。虚线表示的是仅仅基于统一关联参数和列效应所假设的相关关系；实线连接的是完整模型下的期望频数。虚线与实线之间的差别显示了行效应与对角线效应的重要性。

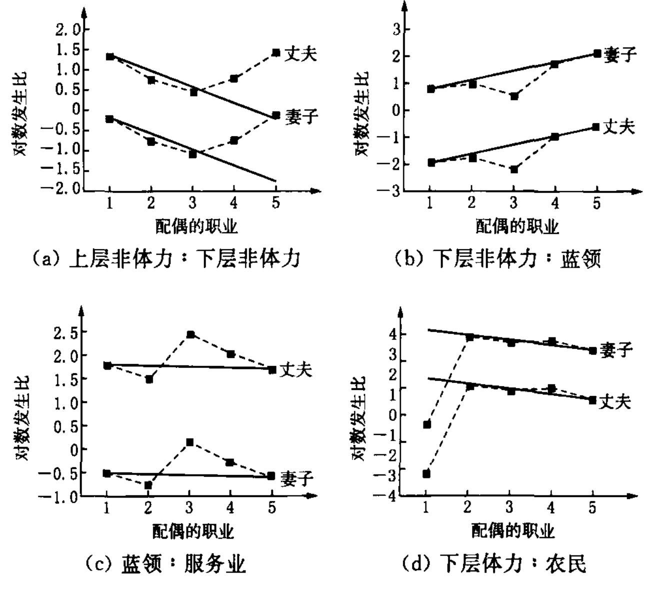


图7.1　同质行与列效应模型（模型Ⅰ）下的期望logit值：丈夫的职业与妻子的职业之间的相关性分析

第 2 节 | 婚姻的宗教匹配

一些研究者可能会认为，表 7.1 中丈夫的职业与妻子的职业的相关性确实体现了婚姻匹配。但是，既然表中的职业是夫妻二人现在的职业，那么，这种相关性既有可能是婚姻匹配所导致的，也有可能是夫妻二人婚后所发生的一系列事件所导致的。然而，如果能够获得夫妻二人的婚前资料，流动模型仍不失为研究婚姻匹配的理想工具。约翰逊(1980)将流动模型应用于婚姻宗教匹配的研究就是一个很好的例子。使用夫妻二人 16 岁时的宗教信仰数据，他建构了第 3 章所介绍的跨越参数模型，以期望获得有关宗教信仰对于婚姻市场之重要性的新见解。

根据前面的介绍，跨越参数模型提出，每相邻两个职业类别之间存在一个跨越界限，其跨越难度由一个参数(v_s)来表示。第(i, j)格上的期望频数是从职业类别 i 到职业类别 j 所需跨越的界限数目的函数。在应用这一模型时，职业类别的排列次序必须是唯一的，只有这样，才能清楚地知道对行类别与列类别的每一个组合来说，究竟跨越了哪些界限。

宗教类别通常来说是没有排列次序的。然而，约翰逊(1980:69—89)根据罗基奇(Rokeach, 1960)所描述的宗教类别之间的“认知上的社会距离”远近提出了一个排序。罗基

奇发现,在美国兰辛和密歇根的两个社区中,浸礼会教徒与天主教徒这两个宗教群体之间在认知上的社会距离最远。他还在研究中发现,如果按照“认知上的社会距离”画一条直线,浸礼会教徒和天主教徒分别位于这条直线的两端,那么其他四个宗教群体——卫理公会教徒、长老会教徒、路德会教徒和圣公会教徒——都可以沿着这个单一尺度进行排序。在进行了一个分类改动(即将长老会教徒与圣公会教徒这两个群体进行合并)之后,约翰逊借助这一排序,用跨越参数模型拟合了宗教群体之间的婚姻匹配数据。但是,美国人的宗教信仰并不仅仅局限于上述六种。其余的宗教倾向(包括“没有宗教信仰”)都被混在一起当做一个剩余分类(而且它包括的人数还相当多),约翰逊将它与作为“主流”宗教群体的基督教徒同等对待。

约翰逊提出了两个模型,具体见表7.2。注意,他将行边缘效应与列边缘效应约束为相等,而且,交互效应具有对称性。两个模型的唯一差别在于对角线的处理。第一个模型(CFS)约束对角线效应,即“同类婚”,对每个宗教类别来说都相同;第二个模型(CS)放松了这一约束。两个模型都包括两类跨越参数,其一是拟合每一个宗教群体与主流宗教群体之间的界限(被设定为不等),其二是拟合除主流宗教群体之外的“其他”宗教群体之间的跨越参数(被设定为相等)。约翰逊分析了四组数据。我只回顾其中一组数据结果,即1973年至1976年的综合社会调查数据(NORC, 1980)。在0.05的显著性水平上,CFS模型能够拟合数据,但CS模型的拟合优度有更显著的提高。模型的参数估计结果见表7.3。表中还报告了由约翰逊计算的两个参数:其一是每个宗教群体人

表 7.2 分析婚姻宗教匹配的 CFS 和 CS 模型:对交互效应的参数化

受访者 16 岁时的宗教信仰	受访者配偶 16 岁时的宗教信仰					
	浸礼会	卫理公会	长老会	路德会	天主教	其他
(1) CFS 模型						
浸礼会	m^*	V_{bm}	$V_{bm}+V_{mp}$	$V_{bm}+V_{mp}+V_{pl}$	$V_{bm}+V_{mp}+V_{pl}+V_{lc}$	V_0
卫理公会	V_{bm}	m^*	V_{mp}	$V_{mp}+V_{pl}$	$V_{mp}+V_{pl}+V_{lc}$	V_0
长老会	$V_{bm}+V_{mp}$	V_{mp}	m^*	V_{pl}	$V_{pl}+V_{lc}$	V_0
路德会	$V_{bm}+V_{mp}+V_{pl}$	$V_{mp}+V_{pl}$	V_{pl}	m^*	V_{lc}	V_0
天主教	$V_{bm}+V_{mp}+V_{pl}+V_{lc}$	$V_{mp}+V_{pl}+V_{lc}$	$V_{pl}+V_{lc}$	V_{lc}	m^*	V_0
其他	V_0	V_0	V_0	V_0	V_0	m^*
(2) CS 模型						
浸礼会	m_b	V_{bm}	$V_{bm}+V_{mp}$	$V_{bm}+V_{mp}+V_{pl}$	$V_{bm}+V_{mp}+V_{pl}+V_{lc}$	V_0
卫理公会	V_{bm}	m_m	V_{mp}	$V_{mp}+V_{pl}$	$V_{mp}+V_{pl}+V_{lc}$	V_0
长老会	$V_{bm}+V_{mp}$	V_{mp}	m_p	V_{pl}	$V_{pl}+V_{lc}$	V_0
路德会	$V_{bm}+V_{mp}+V_{pl}$	$V_{mp}+V_{pl}$	V_{pl}	m_l	V_{lc}	V_0
天主教	$V_{bm}+V_{mp}+V_{pl}+V_{lc}$	$V_{mp}+V_{pl}+V_{lc}$	$V_{pl}+V_{lc}$	V_{lc}	m_c	V_0
其他	V_0	V_0	V_0	V_0	V_0	m_0

注:两个模型都包括对称的边缘效应。

注:两个模型都包括对称的边缘效应。

表 7.3 分析婚姻宗教匹配的 CFS 和 CS 模型:参数估计与比例估计(1973—1976 年)

参数	宗教类别										
	浸礼会		卫理公会		长老会		路德会		天主教		其他
(1) CFS 模型											
边缘效应	0.636		0.154		−0.249		0.047		0.668		—
对角线效应	0.786		0.786		0.786		0.786		0.786		0.786
跨越效应		−0.255		−0.413		−0.389		−0.171		−0.918	
占总人口比例	0.26		0.15		0.09		0.12		0.28		0.10
同类婚比例	0.61		0.42		0.31		0.36		0.62		0.45
(2) CS 模型											
边缘效应	0.303		−0.135		−0.633		−0.565		0.203		—
对角线效应	0.661		0.450		0.750		1.198		0.936		1.944
跨越效应		−0.339		−0.449		−0.363		−0.075		−0.340	
占总人口比例	0.26		0.15		0.09		0.12		0.28		0.10
同类婚比例	0.61		0.37		0.30		0.42		0.62		0.45

注:出于对称性约束的原因,“边缘效应”参数同时适用于行和列。

数占总人口的比例(p_i),其二是模型所预期的每一个宗教群体中选择与本群体成员结婚的人数比例(I_i)。"同类婚"在"其他"宗教群体中表现得最普遍。路德会教徒的同类婚比例也比较高。卫理公会教徒与长老会教徒/圣公会教徒的合并群体之间的距离最远。CS模型与CFS模型之间的主要区别在于对"其他"宗教群体的"同类婚"比例的估计,以及对"其他"宗教群体与主流宗教群体之间社会距离的估计。CFS模型约束"同类婚"比例在每一个宗教群体中都相同,并且将"其他"宗教群体与主流宗教群体之间的社会距离定义得比较高。对这一约束的放松会使"同类婚"比例在"其他"宗教群体中有所增加,作为抵消,改动后的模型对"其他"宗教群体与主流宗教群体间的社会距离的估计会相应降低。

第3节 | 宗教信仰的社会化

麦克雷(McRae，1979)对宗教出身与宗教信仰之间的相关性进行了探索性研究。麦克雷的分析饶有趣味，原因有两个。第一个原因是实质意义上的，很明显，这一研究主题本身令人感兴趣。第二个原因是形式上的，在于分析中引入了其他变量。麦克雷估计了在1958年和1971年的底特律市，父亲和母亲的宗教倾向对于成年子女宗教倾向的影响。他考察了三类宗教群体：新教徒、天主教徒和其他宗教信徒。麦克雷模型的最一般化的形式是分性别的对偶对角线效应模型，它包括三个效应，分别是母亲的影响效应(作用于那些母亲与子女具有相同宗教倾向的单元格)、父亲的影响效应(作用于那些父亲与子女具有相同宗教倾向的单元格)、同类效应(作用于母亲、父亲与子女都具有相同宗教倾向的单元格)。模型的具体表述见表7.4，期望频数和参数估计值见表7.5。请留意，麦克雷对许多参数都进行了约束。

最主要的发现如下：不管是父亲的宗教倾向还是母亲的宗教倾向，都对子女的宗教倾向有显著的影响；父母亲的这种影响在男性样本和女性样本中同样重要；母亲的影响大于父亲的影响；在1958年时，如果父亲和母亲的宗教信仰相同，就会强化儿童信仰“其他”宗教的社会化过程。

表 7.4 宗教信仰的代际传递:麦克雷的对偶对角线效应模型的参数化(相乘形式)

父亲的宗教信仰	母亲的宗教信仰	受访者的宗教信仰 新教	天主教	其他宗教
新教	新教	$d_{pt} m_{pt} h_{pt} q$	1	1
	天主教	d_{pt}	$m_{ct} q$	1
	其他宗教	d_{pt}	1	m_{ot}
天主教	新教	m_{pt}	d_{ct}	1
	天主教	1	$d_{ct} m_{ct} h_{ct}$	1
	其他宗教	1	d_{ct}	m_{ot}
其他宗教	新教	m_{pt}	1	d_{ot}
	天主教	1	m_{ct}	d_{ot}
	其他宗教	1	1	$d_{ot} m_{ot} h_{ot}$

注:该模型同时对两个不同年份时父亲的宗教信仰、母亲的宗教信仰以及受访者的宗教信仰之间的交叉分类数据进行拟合。d 表示父亲的影响效应(给定宗教类别和年份),m 表示母亲的影响效应(给定宗教类别和年份),h 表示同类效应(给定宗教类别和年份)。

表 7.5 宗教信仰的代际传递:麦克雷模型的参数估计值

参数	年份	受访者的宗教信仰 新教	天主教	其他宗教
父亲的影响效应(d)	1958	17	5.9	2.2
	1971	1.7	5.9	2.2
母亲的影响效应(m)	1958	4.3	5.9	2.2
	1971	4.3	5.9	2.2
同类效应(h)	1958	1.0	1.0	7.5
	1971	1.0	1.0	1.0

注:表中所报告的系数是就模型的相乘形式而言的(1.0 表示不显著)。之所以会出现数值相同的系数,是约束条件使然。



总　结

让我来最后总结一下本书对流动表和类似的列联表分析模型的概述。要延续前面概述的研究工作，首先需要将流动表方法和回归方法进行整合。邓肯、布里格、豪特和洛根都在朝这个方向努力，但还有更多的工作要做。其中一项重要的工作便是将这些模型应用于新的研究领域。父母政治倾向对子女政治倾向的影响是相当强的(Knoke, 1976)。夫妻之间的许多共同特征都可以用这些模型来进行研究。追踪研究也同样适用于这种建模方法。

最后，多元变量的建模已经开始成为一个议题。山口、豪特和洛根都考虑到了外在因素对社会流动过程的影响。至于如何将外生变量的限制性效应纳入模型，还有更多的工作需要完成。对社会流动研究来说，这是一个特别重要的领域。希望本书能够推动对上述主题的研究。

参考文献

Baron, J. N. (1980) "Indianapolis and beyond: a structural model of occupational mobility across generations." *American Journal of Sociology 85*: 815—839.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland(1975) *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Blau, P. M., and O. D. Duncan(1967) *The American Occupational Structure*. New York: John Wiley.

Blumen, I., M. Kogan, and P. T. McCarthy(1955) *The Industrial Mobility of Labor as a Probability Process*. Ithaca, NY: Cornell University Press.

Boudon, R. (1975) *Mathematical Structures of Social Mobility*. New York: Elsevier.

Breiger, R. L. (1981) "The social class structure of occupational mobility." *American Journal of Sociology 87*:578—611.

Clogg, C. C. (1983) *Introduction to Structural Models for Qualitative Data*. New York: Academic.

——(1982a) "The analysis of association models for social data." *Journal of the American Statistical Association 77*:803—815.

——(1982b) "Using association models in sociological tables." *American Journal of Sociology 88*:114—134.

——(1981) "Latent structure models for mobility tables." *American Journal of Sociology 86*:836—852.

Davis, J. A. (1974) "Hierarchical models for significance tests in multivariate contingency tables," in H. L. Costner(ed.), *Sociological Methodology, 1973—1974*. San Francisco: Jossey-Bass.

Duncan, O. D. (1981) "Two faces of panel analysis: parallels with comparative cross-sectional analysis and time-lagged association," in S. Leinhardt(ed.), *Sociological Methodology, 1981*. San Francisco: Jossey-Bass.

——(1979) "How destination depends on origin in the occupational mobility table." *American Journal of Sociology 84*:793—803.

——(1978) Personal communication.

——(1966) "Methodological issues in the analysis of social mobility," in N.

J. Smelser and S. M. Lipset(eds.), *Social Structure and Economic Development*. Chicago: Aldine.

——(1961) "A socioeconomic index for all occupations," in A. J. Reiss, Jr. (ed.), *Occupations and Social Structure*. Glencoe, IL: Free Press.

——and H. Schuman(1980) "Effects of question wording and context: an experiment with religious indicators." *Journal of the American Statistical Association 75*:269—275.

Featherman, D. L., and R. M. Hauser(1978) *Opportunity and Change*. New York: Academic.

——M. Sobel, and P. Dickens(1975) "A manual for coding occupations." Center for Demography and Ecology, Paper 75—1, University of Wisconsin.

Fienberg, S. E. (1980) *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA: MIT Press.

Glass, D. V. (1954) *Social Mobility in Britain*. London: Routledge & Kegan Paul.

Goldberger, A. S. (1973) "Structural equation models," in A. S. Goldberger and O. D. Duncan(eds.), *Structural Equation Models in the Social Sciences*. New York: Seminar.

Goldthorpe, J. W., and K. Hope(1971) "Occupational grading and occupational prestige," in K. Hope(ed.), *The analysis of Social Mobility: Methods and Approaches*. Oxford: Clarendon.

Goodman, L. A. (1981) "Criteria for determining whether certain categories in a cross-classification table should be combined." *American Journal of Sociology 87*:612—650.

——(1979a) "Simple models for the analysis of association in cross-classifications having ordered categories." *Journal of the American Statistical Association 74*:537—552.

——(1979b) "Multiplicative models of the analysis of mobility tables and other kinds of cross-classification tables." *American Journal of Sociology 84*:804—819.

——(1974) "The analysis of systems of qualitative variables when some of the variables are unobservable." *American Journal of Sociology 79*: 1179—1259.

——(1972a) "Some multiplicative models for the analysis of cross-classified data," in J. Neymann(ed.), *Proceedings of the Sixth Berkeley Sympo-*

sium on Mathematical Statistics and Probability. Berkeley: University of California Press.

——(1972b) "A modified regression approach to the analysis of dichotomous variables." *American Sociological Review 37*:28—46.

——(1972c) "A general model for the analysis of surveys." *American Journal of Sociology 77*:1035—1086.

——(1969a) "How to ransack social mobility tables and other kinds of cross-classification tables." *American Journal of Sociology 75*:1—39.

——(1969b) "On the measurement of social mobility: An index of status persistence." *American Sociological Review 34*:831—850.

——(1968) "The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables with or without missing entries." *Journal of the American Statistical Association 63*: 1091—1131.

——(1965) "On the statistical analysis of mobility tables." *American Journal of Sociology 70*:564—585.

——(1961) "Statistical methods for the mover-stayer model." *Journal of the American Statistical Association 56*:841—868.

Haberman, S. J. (1979) *Analysis of Qualitative Data*. New York: Academic.

——(1974) "Log-linear models for frequency tables with ordered classifications." *Biometrics 30*:589—600.

Hauser, R. M. (1981) "Hope for the mobility ratio." *Social Forces 60*: 572—584.

——(1979) "Some exploratory methods for modeling mobility tables and other cross-classified data," in D. R. Heise(ed.), *Sociological Methodology, 1980*. San Francisco: Jossey-Bass.

——(1978) "A structural model of the mobility table." *Social Forces 56*: 919—953.

——P. J. Dickinson, H. P. Travis, and J. M. Koffel(1975a) "Structural changes in occupational mobility among men in the United States." *American Sociological Review 40*:585—598.

——(1975b) "Temporal changes in occupational mobility: evidence for men in the United States." *American Sociological Review 40*:279—297.

Hodge, R. W. (1981) "The measurement of occupational status." *Social Science Research 10*:396—415.

Hout, M. (forthcoming) "Status, autonomy, and training in occupational mobility."

——(1982) "The association between husbands' and wives' occupations in two-earner families." *American Journal of Sociology 88*:397—409.

Johnson, R. A. (1980) *Religious Assortative Marriage in the United States*. New York: Academic.

Knoke, D. (1976) *Change and Continuity in American Politics*. Baltimore: Johns Hopkins University Press.

——and P. J. Burke(1980) *Log-Linear Models*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07—020. Beverly Hills, CA: Sage.

Lazarsfeld, P. F., and N. W. Henry(1968) *Latent Structure Analysis*. Boston: Houghton-Mifflin.

Lenski, G. (1966) *Power and Privilege*. New York: McGraw-Hill.

Lipset, S. M., and R. Bendix(1959) *Mobility in Industrial Society*. Berkeley: University of California Press.

——(1952) "Social mobility and career patterns." *American Journal of Sociology 57*:494—504.

Logan, J. A. (1983) "A multivariate model for the mobility table." *American Journal of Sociology 87*.

Macdonald, K. I. (1981) "On the formulation of a structural model of the mobility table." *Social Forces 60*:557—571.

McRae, J. A. (1979) "The stability of religious differences in primary group attitudes and behaviors." Ph. D. dissertation, University of Arizona.

National Opinion Research Center [NORC](1980) *General Social Surveys: Cumulative Codebook*. Chicago: National Opinion Research Center and Roper Center.

Pontinen, S. (1981) "Models and social mobility research: a comparison of some log-linear models for a social mobility matrix." Department of Sociology, University of Helsinki. (mimeo)

Pullum, T. (1975) *Measuring Occupational Inheritance*. New York: Elsevier.

Rogoff, N. (1953) *Recent Trends in Occupational mobility*. Glencoe, IL: Free Press.

Rokeach, M. (1960) *The Open and Closed Mind*. New York: Basic Books.

Simon, G. (1974) "Alternative analyses for the singly-ordered contingency table." *Journal of the American Statistical Association 69*:971—976.

Singer, B., and S. Spilerman(1976) "The representation of social processes by Markov models." *American Journal of Sociology 82*:1—54.

——(1974) "Social mobility processes for heterogeneous populations," in H. L. Costner (ed.), *Sociological Methodology, 1973—1974*. San Francisco: Jossey-Bass.

Svalastoga, K. (1959) *Prestige, Class, and Society*. Copenhagen: Gyldendal.

Tyree, A. (1973) "Mobility ratios and association in mobility tables." *Population Studies 27*:577—588.

U. S. Department of Labor (1977) *Dictionary of Occupational Titles: Fourth Edition*. Washington: Government Printing Office.

White, H. C. (1963) "Cause and effect in social mobility tables." *Behavioral Science 7*:14—27.

Wright, E. O., and L. Perrone(1977) "Marxist class categories and income inequality." *American Sociological Review 42*:32—55.

Yamaguchi, K. (1983) "The structure of intergenerational occupational mobility: generality and specificity in resources, channels, and barriers." *American Journal of Sociology 88*:718—745.

译名对照表

additive property	可叠加性
aggregation	集合
assortative marriage	婚姻匹配
baseline level	基准层级
baseline model	基准模型
basic set	基本单元
Bureau of the Census	人口普查局
class barrier	阶级隔阂
cluster	聚类
cognitive social distances	认知上的社会距离
column effect	列效应
column total	列边缘和
conditional distribution	条件分布
conditional probability	条件概率
constancy	恒定性
constraint	约束
covariance matrix	协方差矩阵
cross-classified data	列联表数据
cross-national constraint	跨国家约束
cross-products ratio	叉积比
Current Population Survey	人口现状调查
degrees of freedom	自由度
density	强度
design matrix	设计矩阵
destination	职业获得/地位获得
Dictionary of Occupational Titles	职业名称词典
disinheritance	非继承性
endogamy	同类婚
equality constraint	等值约束
equivalence	等价性
equivalent	等价的

equivalent expressions	等价方程式
expected logit	期望 logit 值
expected odds-ratio	期望优比
generalization	一般化形式
General Social Survey	综合社会调查
goodness of fit	拟合优度
grand mean	总均值
immobility effect	继承效应
imply	蕴含
indeterminate results	不确定结果
index of dissimilarity	相异指数
index of social distance	社会距离指标
inflow	流入
interaction effect	交互效应
intergenerational mobility	代际流动
intragenerational mobility	代内流动
invariance under transformation	变形中的不变性
iterative method	迭代方法
joint density	联合强度
latent classes	潜在类别
latent social classes	潜在的社会阶级
log-odds	对数发生比
log-odds ratio	对数优比值
main diagonal	主对角线
marginal distribution	边缘分布
marginal effect	边缘效应
marginal parameter	边缘效应参数
marginals/marginal total	边缘分布频数
minor diagonal	次对角线
mobility ratio	流动比率
mobility ratios	流动比率
mobility status	流动状态

multiplicative form	相乘形式
nonstayer	移动者
normalization	标准化方法
null hypothesis	零假设
oblique	嵌入的
observed frequency	观测频数
occupational status	职业地位
odds	发生比
odds ratios	优比
origin	职业出身/社会出身
outflow	流出
overfitting	过度拟合
panel data	面板数据
panel study	追踪研究
parameter	参数
partition	划割
path analysis	路径分析
persistence	持续性
prevalence effect	普遍效应
quasi-independent	准独立的
Quasi-Perfect Mobility Model(QPM)	准完全流动模型
random error	随机误差
religious origins and destinations	宗教出身与宗教信仰
row effect	行效应
row total	行边缘和
restriction	条件限定
scale	测量尺度
scale adjustment	尺度调整参数
scaling	度量
scoring index	得分尺度
simultaneous inference	互相参照
socioeconomic status	社会经济地位

square table	对称表
statistic	统计量
status barrier	阶层隔阂
stayer	滞留者
stepwise regression	逐步回归
subset	子集
systematic association	系统性的相关
tabulation of the χ^2 distribution	卡方分布表
unbiased estimate	无偏估计
uniform association parameter	统一关联参数
very insensitivity	不变性
zero constraint	零值约束
zero-sum normalization	零和规一
association model	关联模型
asymmetrical model	非对称模型
constrained diagonals model	对角线约束模型
corners model	边角模型
Crossing Parameters Model(CP)	跨越参数模型
diagnostic model	诊断模型
Diagonal Row-Column Model(DRC)	对角线行—列模型
diagonals model	对角线模型
Diagonals(Constrained) Model(D-C)	对角线(约束)模型
Diagonals Crossing Model(DC)	跨对角线模型
equivalent model	等价模型
general model	一般化模型
hierarchical model	层级模型
homogeneous row and column effects model(Model Ⅰ)	同质行与列效应模型(模型Ⅰ)
latent structure model	潜在结构模型
linear-by-linear interaction model	线性乘线性交互模型
Marginal Homogeneity Model(MH)	边缘齐性模型
modified crossing parameters model	跨越参数改动模型

mover-stayer model	移动者—滞留者模型
multiplicative model	相乘模型
null association model	零相关模型
pairwise diagonal effects model	对偶对角线效应模型
perfect mobility model	完全流动模型
quasi-column effects model	准列效应模型
quasi-crossing parameters model	准跨越参数模型
quasi-homogenous row and column effects model	准同质行与列效应模型
quasi-latent structure model	准潜在结构模型
quasi-row effects model	准行效应模型
Quasi-Symmetry Model(QS)	准对称模型
quasi-uniform association model	准统一关联模型
row and column effects model(Model Ⅰ)	行与列效应模型(模型Ⅰ)
social distance model	社会距离模型
standard latent structure model	标准潜在结构模型
Status, Autonomy, and Training Model (SAT)	地位、自主性和培训模型
Symmetrical Diagonals(Constrained) Model(SD-C)	对角线对称(约束)模型
symmetry model	对称模型
topological model	拓扑模型
unconstrained diagonals model	对角线非约束模型